DOCUMENT RESUME

ED 308 231 TM 013 611

AUTHOR Wilder, Gita Z.; Powell, Kristin
TITLE Sex Differences in Test Performance: A Survey of the Literature. College Board Report No. 89-3.
INSTITUTION College Entrance Examination Board, New York, N.Y.
REPORT NO ETS-RR-89-4
PUB DATE 89
NOTE 59p.
AVAILABLE FROM College Board Publications, Box 886, New York, NY 10101 ($6.00).
PUB TYPE Information Analyses (070)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Academic Achievement; *Achievement Tests; Biological Influences; *College Entrance Examinations; Demography; Elementary Secondary Education; Higher Education; Literature Reviews; Longitudinal Studies; Mathematics Tests; National Surveys; *Sex Differences; Social Psychology; *Standardized Tests; *Test Bias; Testing Programs; Test Validity; Verbal Tests
IDENTIFIERS High School and Beyond (NCES); National Assessment of Educational Progress; National Longitudinal Study High School Class 1972; Scholastic Aptitude Test

ABSTRACT

A review of the significant research on gender differences in test performance is provided, including studies documenting these differences and theories forwarded for the factors related to differential test performance. Research covered addresses undergraduate admissions tests, graduate and professional school admission tests, validity studies, national studies, verbal ability tests, quantitative ability tests, and voluntary testing programs. Specific testing programs discussed include the National Assessment of Educational Progress, National Longitudinal Study of 1972, High School and Beyond study, and Scholastic Aptitude Test. Efforts at explaining performance differences include biological, social, psychological, demographic, and educational factors as well as test characteristics. In so far as there is convergence among studies, it appears that the disparities between the sexes are slowly diminishing. Males have caught up with females in tests of verbal ability and achievement, while females have narrowed the gap between themselves and their male counterparts in the area of mathematics achievement. Concern continues over social implications of the slight differences in performance. A 235-item list of references and six flowcharts are provided. A bibliography of pertinent references, arranged by format and topic, is appended. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308231

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

ROBERT G. CAMERON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

College Board Report

No. 89-3

Sex Differences in Test Performance: A Survey of the Literature

Gita Z. Wilder
Kristin Powell

Sex Differences in Test Performance: A Survey of the Literature

Gita Z. Wilder
Kristin Powell

With a Foreword by Gretchen Rigol

College Board Report No. 89-3
ETS RR No. 89-4

College Entrance Examination Board, New York, 1989

Gita Wilder is a research scientist at Educational Testing Service, Princeton, New Jersey.

Kristin Powell is a research assistant at Educational Testing Service, Princeton, New Jersey.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

Figure 1. Reprinted by permission of University of Nebraska Press from J.S. Eccles, 1984. Sex differences is achievement patterns. Figure 1. In T. Sonderegger, ed., *Nebraska Symposium on Motivation.* Copyright 1985 by the University of Nebraska Press.
Figure 2. Reprinted by permission of Educational Testing Service from M. Lockheed, et al. 1985. *Sex and ethnic differences in middle school mathematics, science, and computer science: What do we know?* Figure 4.
Figure 3. Reprinted by permission of Jai Press, Inc., from A. Grieb and J. Easley, 1984. A primary school impediment to mathematical equity: Case studies in rule-dependent socialization. Figure 2. In M.W. Steinkamp and M.L. Maehr, eds., *Advances in Motivation and Achievement.*
Figure 4. Reprinted by permission from C.A. Ethington and L.M. Wolfle, 1986. A structural model of mathematics achievement for men and women. Figure 1. *American Educational Research Journal.* Copyright 1986 by the American Educational Research Association (Washington).
Figure 5. Reprinted by permission of Jai Press, Inc., from S. Kavrell and A. Petersen, 1984. Patterns of achievement in early adolescence. Figure 2. In M.W. Steinkamp and M.L. Maehr, eds., *Advances in Motivation and Achievement.*
Figure 6. Reprinted by permission of Lawrence Erlbaum Associates, Inc., from L.L. Wise, 1985. Project TALENT: Mathematics course participation in the 1960s and its career consequences. Figure 2.4. In S.F. Chipman, et al., eds., *Women and Mathematics: Balancing the Equation.*

The College Board is a nonprofit membership organization committed to maintaining academic standards and broadening access to higher education. Its more than 2,600 members include colleges and universities, secondary schools, university and school systems, and education associations and agencies. Representatives of the members elect the Board of Trustees and serve on committees and councils that advise the College Board on the guidance and placement, testing and assessment, and financial aid services it provides to students and educational institutions.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101. The price is $6.

Copyright © 1989 by College Entrance Examination Board. All rights reserved.
College Board, Advanced Placement Program, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

CONTENTS

	Page
Foreword	v
Introduction	1
The Data Concerning Gender Differences	2
Undergraduate Admission Tests	2
Graduate and Professional School Admission Tests	3
Data from Validity Studies	3
Tests Involving Nationally Representative Samples	4
Verbal Ability	6
Quantitative Abilities	7
Trends in Sex Differences in Performance	9
Voluntary Testing Programs	9
Nationally Representative Samples	10
Are Gender Differences Disappearing?	14
Efforts at Explanation	14
Biological Explanations	15
Social and Psychological Explanations	16
Individual Differences	18
Educational Variables	19
Integrative Models	22
Demographic Explanations of Trends	23
Characteristics of the Tests Themselves	25
Summary	28
Discussion	30
References	32
Appendix A: References Arranged by Format and Topic	39
Appendix B: Selected Models of Influences on Gender-Based Differences	46

Figures

1. Reduced path-analytic diagram for test of socialization model	46
2. Task-performance model of mathematics, science, or computer performance	47
3. Alternative pathways of mathematical development	48
4. Structural equation and measurement models of mathematics achievement	48
5. A model of biopsychosocial influences on cognitive performance	49
6. Summary path model of the relationship of sex to high school mathematics achievement	50

Tables

1. Average Writing Achievement (ARM): Nation, Males, and Females 5

2. NLS/HS&B IRT Mean Scaled Scores (High School Seniors) 5

3. National Assessment of Educational Progress Literacy Levels for Young Adults (21–25 Years Old), 1985 7

4. Mean LSAT Scores and GPA, by Sex 10

5. Mean MCAT Scores of Male and Female Medical School Applicants 10

6. National Assessment of Educational Progress Trends in Mean Reading Proficiency for the Nation, Males, and Females 11

7. National Assessment of Educational Progress Trends in Male/Female Differences on Three Writing Tasks 12

8. Trends in Average Mathematics Proficiency for 9-, 13-, and 17-Year-Olds by Gender . . 12

9. Changes in NLS/HS&B Mean Test Scores (High School Seniors) 13

10. Changes in SAT and NLS-HS&B Test Scores, 1972–1982 13

FOREWORD

As is evident from a cursory review of the Reference section of this report, there has been extensive research during the past several decades that documents and attempts to explain and understand the differences between men and women on a wide range of educational outcomes. Although educators and researchers have long been aware that such differences exist, it is only recently that public attention has focused on the topic. Such scrutiny is welcome and appropriate; however, it is disappointing that many of the articles that have appeared in the popular press focus rather narrowly on only a few aspects of the issue. Even more unfortunate is the lack of understanding of the pervasiveness of these differences and the complex factors that might be contributing to the differences.

This report, therefore, represents a timely and useful summary of significant research that has already been conducted and provides a context for future evaluation. More importantly, it discusses various hypotheses that have been advanced to explain observed differences and suggests interventions that might work toward eliminating them.

As the authors note in the introduction, "the conclusions about gender differences that can be reached at the current time are limited" and that "the data that support many of the contentions made about gender differences and their cause are inconclusive and often contradictory." Nonetheless, this comprehensive review of the literature supports several generalizations about standardized tests.

- Many different tests given over a wide range of ages and educational levels reveal male-female score differences.
- In general, the largest differences appear in tests of mathematical or quantitative ability, where men tend to do better than women, particularly in secondary school and beyond. In recent years, there is some evidence that this gap may be narrowing.
- Women have tended to do slightly better than men in many tests of verbal skills (particularly in writing), but a number of studies have shown that this superiority has diminished since the early 1970s.

This report also contains a valuable summary of explanations that have been advanced about these observed differences, ranging from theories related to hormonal and other biological causes to differences in social and educational backgrounds. For those of us who are not statisticians or researchers, the schematic integrative models included in Appendix B may seem overwhelming, but the message that there are probably a very complex series of factors underlying male/female differences seems both reasonable and compelling.

The reader is invited to weigh the plausibility and relative merit of each of the individual and integrative hypotheses. The explanations that relate to educational and social experiences probably warrant the most attention since these are aspects of a young person's life that we can do something about. For example, the authors document numerous studies that reveal the different ways girls and boys are treated at home and in classrooms, as well as the more subtle messages conveyed through books, televisions, and other experiences about acceptable or expected attitudes and behavior. Some of these experiences are more easily modified than others. Encouraging or requiring young women to enroll in more advanced mathematics and science courses is relatively straightforward, but changing attitudes toward the importance and relevance (and general interest in) these subjects is considerably more complex.

The SAT is perhaps the most frequently cited test in discussions of various gender issues. Indeed, there is such a wealth of data available about the SAT and the students who take it that many researchers have utilized these data in numerous studies. The average SAT scores for men and women have hardly ever been identical since the SAT was first administered in 1926. For many years, the average scores for men were higher in the mathematical sections of the test and women earned higher average verbal scores. People tended to accept these differences as the normal state of affairs and there was little, if any, public discussion about the topic. In 1972, however, the average verbal score for women fell slightly below that of men and has remained below by about 10 points for the past decade. It should be noted that the differential in SAT-math has remained at 40 to 50 points for more than two decades, possibly longer.

A great deal of attention has been devoted to understanding these differences. Part of the explanation relates to the fact that the SAT-taking population is a self-selected group and that the backgrounds of the women who choose to take the SAT are, on average, different from the backgrounds of the men taking the test. Perhaps the most obvious difference is sheer numbers: there are now considerably more women than men taking the SAT (46,000 more in 1988). In addition, the women taking the test are less likely than men to have completed as many advanced college-preparatory courses (particularly in mathematics and science). There are also differences in other background characteristics, such as the fact that women are much more likely to come from families where neither parent attended college, suggesting differences in their home environments. Research cited in this survey of the literature sug-

gests that the differences in the average verbal scores are eliminated if these background characteristics are controlled. Similarly, about half of the difference in the average math scores are accounted for by background differences.

There is the temptation to explain away differences by simply dismissing the instrument as "biased." this survey cites 36 different studies related to the characteristics of the tests and discusses a number of the more recent studies that have been conducted on the SAT and other standardized tests. Although occasionally, individual test questions are identified that seem to be differentially easier or more difficult for one group or another, there appears to be no clear pattern nor simple explanation for these differences. Current test development procedures for the SAT require numerous subjective reviews and statistical checks to assure that questions that might give an unintended advantage or disadvantage to either males or females are not included in the test. Most other test sponsors utilize similar procedures, yet differences continue to be evident in nearly all standardized tests.

One of the most baffling aspects of this issue is that women's scores on the SAT and many other tests are often lower than men's, but women tend to receive higher grades in both high school and college. Further confounding this issue is that the SAT and other admission test scores relate more closely to college grades for women than for men. One of the most widely accepted definitions of an appropriate and fair test is that the results must work equally well for all subgroups of the population. Studies of predictive validity show that the average correlation between admission scores and college grades is actually higher for women than men at most colleges. At the same time, however, colleges that use a combined male/female prediction formula will often see slight "under-prediction" for females because of the higher overall grades that women tend to earn. The authors note that this may be related to differences in course selection and/or grading practices. Indeed, women tend to concentrate their studies (in both high school and college) in the humanities, where higher average grades are the norm. Men, on the other hand, tend to gravitate toward subjects such as math and science that, on average, seem to have more rigorous grading standards. Relatively little research has been conducted about differential grades, but just as objective test results have been scrutinized, so should grades also be examined. What does and should an "A" represent? To what extent does it reflect mastery of the subject, punctuality, work habits, attentiveness, etc.?

Clearly, these are very complex issues and there are still many unanswered questions about sex differences in several academic contexts. This literature survey should engender further research that may lead us to find ways to eradicate the differences that can be observed in practically every standardized measure we know. The young women in our country deserve the same educational, economic, and social opportunities that are open to men. Just as the well-publicized SAT score decline influenced the school reform movement, the widespread awareness of score differences between men and women can direct us toward some positive initiatives to make equal opportunity a reality for us all.

GRETCHEN WYCKOFF RIGOL
Executive Director
Access Services

INTRODUCTION

Although gender differences in cognitive functioning, and theories about their origins, have been with us at least since the turn of the century, scholars in the area generally credit the appearance in 1974 of Eleanor Maccoby and Carol Nagy Jacklin's book, *The Psychology of Sex Differences*, with providing the best starting point for any data-based discussion of the phenomenon. In their summary and evaluation of a large body of work on sex differences, Maccoby and Jacklin repudiated a number of common claims for sex differences, left a number (mostly affective and personality characteristics) open to future discussion, and identified four "fairly well established" differences. Their review took the form of a "head count" of studies that examined sex differences in various domains. The four arenas in which Maccoby and Jacklin acknowledged documented sex differences were verbal ability, in which females seemed to excel; and visual-spatial ability, mathematical ability, and aggression, all favoring males. Critics of their conclusions point to large disparities among the studies included and the relatively small size of many of the differences cited.

More recently, researchers have employed meta-analytic techniques to examine the phenomenon of sex differences in several of the areas identified by Maccoby and Jacklin (Linn and Hyde 1986; Hyde and Linn 1988). Meta-analytic techniques go beyond a simple count of studies with particular findings and concentrate, instead, on estimating the size of the effect in question across a number of studies that meet specified criteria. While such techniques offer a more systematic approach to aggregating results across disparate studies, they have generated new controversies related to the legitimacy of aggregating across studies of varying quality and the significance of the effects they document (e.g., Chipman 1988; Rosenthal and Rubin 1982).

During the past decade, concern about sex differences in measured performance has taken a new turn. For many years, performance on standardized tests seemed to verify the 1974 conclusions of Maccoby and Jacklin in that males consistently outperformed females on tests of mathematical and spatial ability, and females outperformed males on tests of verbal ability. Research in the area directed itself to the possible reasons for the differences. In recent years, however, women have lost their relative advantage in the verbal area, at least where undergraduate admission tests have been concerned. Since 1972, women seem to have performed progressively less well relative to men on verbal measures, whereas men's performance has improved. Meanwhile, women's relative disadvantage on mathematical and spatial tasks seems to have remained constant or diminished, depending on whose assessment is invoked.

The purpose of this review is to examine the current data on sex differences in test performance along with some of the hypotheses and evidence concerning possible causes of the differences. At the outset it should be noted that the conclusions about gender differences that can be reached at the current time are limited. For all the attention that the subject has received, the data that support many of the contentions made about gender differences and their causes are inconclusive and often contradictory. The majority of studies lack generalizability, based as they are on different populations or on performance in limited domains by small samples of individuals. While the techniques of meta-analysis permit aggregation of data across such studies, there continues to be debate over the interpretation of effect sizes calculated in such analyses (Cohen 1969; Rosenthal and Rubin 1982). One important source of information from relatively large samples is the results of tests for admission (to undergraduate, graduate, or professional schools) that are simply not representative of the general population. Any examination of trends in these data is complicated by changes over time in the nature of the population of candidates for admission to the educational programs in question, and possibly in the content of the tests. Data from large-scale studies based on nationally representative samples (the National Assessment of Educational Progress, for example) avoid the self-selection bias inherent in the admission test data but do not lend themselves to the kinds of fine-grained analyses permitted by small-sample studies. The nature of the evidence is clearly an important mediator of the kinds of conclusions that can be derived. Complicating the issue still further are the different conclusions that researchers have managed to reach even when they work from the same data.

This review is organized into three major sections. This first examines the data, evidence (or lack thereof) for differential performance by males and females on various tests. The second section examines a variety of possible correlates and hypothesized causes of the reported differences. These include demographic and social trends, individual differences that span the range from biological to psychosocial, and characteristics of the tests themselves. The final section identifies several areas for continuing or future research.

One final note: Any review of a body of material as large as the work on gender differences must be selective. Not only must some studies be left out while others are included; some are treated in greater detail than others. In selecting studies for inclusion, as well as in choosing among studies for more intensive treatment within the review, the emphasis has

been on recent research, on studies and reviews conducted after 1980. Where work conducted before 1980 is the most recent, it is included.

THE DATA CONCERNING GENDER DIFFERENCES

This section of the review will consider some of the evidence concerning sex differences in test performance. Although there is no simple way to organize such an overview, this one will start with some of the more general findings from large-scale data bases and then move to separate consideration of the verbal and quantitative domains.

Undergraduate Admission Tests

One of the major sources of concern in the area of gender differences in test performance is the growing realization that such differences may exercise a negative effect on the educational opportunities of one group (more frequently women) or another. Nowhere is that concern more evident than in the admission process that governs access to undergraduate, graduate, and professional schools. Of the measures used to guide those who make decisions about candidates, the Scholastic Aptitude Test (SAT) has garnered the most critical attention in recent years for its impact not only on women but on various racial and ethnic minority groups. The American College Testing Program Examination (ACT) is also taken each year by large numbers of high school juniors and seniors seeking admission to college.

The SAT is described as "a measure of developed abilities" (Donlon 1984) and produces separate scores for verbal (SAT-verbal) and mathematical (SAT-mathematical) subsections and the Test of Standard Written English (TSWE). In the SAT population, the average mathematical score difference has been about half of a standard deviation in favor of males for most of the years since the SAT was introduced. By way of contrast, and consistent with the findings of Maccoby and Jacklin, women's average SAT-verbal scores tended to be slightly higher than men's until the late 1960s. At that point a downward trend in women's scores began. By 1980 women's average verbal score was 12 points below men's, a difference of about .11 standard deviation. (The trends in these data will be considered in a later section of this review.)

Four subsections make up the ACT: English usage, mathematics usage, social studies reading, and natural science reading. The last two subsections combine items that measure reading comprehension and items based solely on prior knowledge of subject matter in proportions, respectively, of 70 and 30 percent (Burton 1987). Dauber (1987) examined gender differences in performance on the various subsections of these two tests among students who took the tests in 1984–85 and 1985–86, and computed the significance of the effects they document (after Cohen 1977) to assess the magnitude of the observed differences[1] The largest effect sizes were found for SAT-mathematical scores (.41 standard deviation), ACT natural science reading (.40), ACT mathematics usage (.34), and ACT social studies reading (.23), all favoring males. Cohen (1977) labels these effect sizes "small," although Dauber underscores the practical significance of the differences by calling attention to the fact that, for example, the ratio of males to females who scored at the ninetieth percentile for SAT-mathematical sections was 2.6:1.

Effect sizes for both the TSWE and ACT English were quite small (.12 and .16 respectively) and favored females. The effect size was even smaller (.10) for the SAT-verbal sections and favored males. Dauber labels all three differences "slight" and claims little practical significance for them.

Stanley (1987) examined the differences between males' and females' performance on the College Board Achievement Tests in 1982, 1983, 1984, and 1985, and Advanced Placement Examinations in 1984, 1985, and 1986; he found differences that were somewhat larger for the Achievement Tests than for the Advanced Placement Examinations. The analysis of Achievement Test data yielded 56 effect sizes for 14 different tests. Averaging across the four years, Stanley found that females scored higher than males on four of the 14 tests: English Composition, German, Hebrew, and Literature. All four of the effects measured were small (.10 standard deviation or less), negligible, in fact, according to Cohen's (1977) classification. Differences favoring males were found for all the remaining tests, such differences ranging in magnitude from .59 for Physics to .06 for Spanish. "Moderate" differences were found for European History (.58) as well as for Physics; "small" differences were found for American History (.40), Chemistry (.39), Mathematics Level I (.39), Mathematics Level II (.38), Biology (.36), and Latin (.20). The remaining differences that favored males were negligible. The data also showed that males tended to score higher, and higher relative to females, as their representation among test-takers increased. That is, the highest effect sizes favoring males occurred

1. There are a number of ways of computing effect size, but all have in common the creation of a standardized measure that is expressed in terms of standard deviation units. In Dauber's case, the computation involved dividing the mean difference between males and females on a given test by the square root of the mean of the two variances (Dauber, p. 2).

for tests that high proportions of males took. For example, 81 percent of the Physics test-takers and 65 percent of the European History test-takers were male. By way of contrast, 72 percent of those who took the French test during the period in question were female.

Analysis of the Advanced Placement data yielded 72 different effect sizes, which were reduced by averaging the results for each test across the three years. The largest mean effect size (.50) was for Computer Science, an examination for which the test population was 85 percent male. Although the other effect sizes were smaller, the five largest (Physics B, .41; Chemistry, .33; Physics C: Mechanics, .37; Physics C: Electricity and Magnetism, .29; in addition to Computer Science) were for tests taken by the most males. Stanley points out that whereas the effect size favoring males in Computer Science declined between 1983 and 1985 from .59 to .37, the percentage of males (85) taking the test remained constant.

Males outperformed females in five additional AP examinations by effect sizes considered small: Music Listening and Literature, .27; European History, .26; American History, .24; Biology, .22; and Calculus AB, .20. There were altogether seven main effect sizes that favored females, none larger than .15 for Latin. In this test, effect sizes for females declined over the three years; gains in the performance of females relative to males were found for AP examinations in American History, Studio Art, Computer Science (as noted above), French, Music Theory, and Physics B, although males still outperformed females to a moderate extent in the latter.

Stanley warns, as a result of his analysis, that the gender disproportions among test-takers combine with the differences in performance to place women at a disadvantage where undergraduate admission and advanced placement and credit are concerned, especially in science, mathematics, and history.

Graduate and Professional School Admission Tests

Test results also contribute to decisions about admissibility to graduate and professional schools. The Medical College Admission Test (MCAT) is intended to evaluate applicants' understanding of concepts in biology, chemistry, and physics and to assess their analytical abilities in the context of problems and data that have some relevance to the field of medicine. There are four parts to the test, and the results are reported in terms of six scores: Biology, Chemistry, Physics, Science Problems, Reading Skills Analysis, and Quantitative Skills Analysis. Results from the Graduate Management Admissions Test (GMAT) are used by about 850 graduate programs in management to guide admission decisions. The test results are reported as Total, Quantitative, and Verbal scores. The Graduate Record Examinations (GRE), used by graduate schools in their admissions process, include both a General Test and Subject Tests. The latter test the knowledge and understanding attained in the course of an undergraduate major concentration in a given field. The General Test includes sections that measure verbal, quantitative, and analytical abilities. Finally, the Law School Admissions Test (LSAT) includes sections that measure reading comprehension, analytical reasoning, evaluation of facts and logical reasoning, and a writing sample which is not scored.

Brody (1987) examined gender differences in performance on these tests in various administrations between 1980 and 1985, and compared the mean results for males and females across the tests. She concluded that, for aptitude tests, the greatest differences were on the quantitative measures, i.e., the GRE-Q, the GMAT-Q, and, to a lesser extent, the MCAT-Q, all of which favored males. The latter finding corroborates an earlier set of findings concerning score differences on the MCAT by Jones (1984), who observed that women scored lower than men on five of the six subtests and concluded that "the rank ordering of these differences parallels the degree of quantitativeness of each MCAT subtest." Specifically, he reported reading scores for male and female applicants that were roughly equivalent; biology scores that were about one-half point greater for men; chemistry, science problems, and quantitative scores about one full point greater for men; and physics scores about one-and-one-half points greater for men.

Brody found no differences on the verbal portions of the GRE and on the LSAT, which is largely verbal; a difference on the GMAT-verbal favored females. Tests of subject matter knowledge, i.e., the GRE subject tests and (as noted above) the science subtests of the MCAT, favored males. The greatest differences were in mathematics, the physical sciences, political science, and history.

Data obtained directly from the Law School Admissions Service about the LSAT show virtually identical scores for males and females, although the scores for female test-takers are slightly lower than the men's—a fraction of a whole point on a 48-point scale—whereas their grade-point averages are slightly higher—3.10 compared with 2.99 in 1986–87 (Christensen 1988).

Data from Validity Studies

Additional evidence concerning sex differences in test performance comes from validity studies of admission tests, which tend to compare test scores with some criterion, usually first-year grade-point

average. For example, combining data for 685 colleges for which SAT scores, high school records, and freshman grade-point average were available separately for males and females, Ramist (1984) reported that women's grade-point average was better predicted than men's. In these data, which covered the years 1964 to 1981, the SAT correlation, high school record correlation, and SAT and high school record multiple correlation were higher for women than for men. Because women traditionally have higher freshman grade-point averages than men (Wild 1977), the use of a single regression equation for both sexes results in underprediction for women and overprediction for men. An examination of trends in the data over the period in question led Ramist to conclude that the male–female differences in the correlations were getting smaller. Linn (1982) summarized studies using the SAT, the ACT, and the LSAT, and reported a tendency for the test scores of females to correlate more strongly than males' with performance measures and therefore to predict subsequent academic performance more accurately. In cases where accuracy of prediction is similar, Linn corroborated the Donlon finding that test scores have systematically underpredicted the performance of females. Jones and Vanyur (1985) examined scores on the six areas assessed by the MCAT and found underprediction of females' grade-point averages in the second year of medical school when the chemistry and physics scores (but not the remaining measures) were used in a single prediction equation for both sexes. Breland and Griswold (1982) found that women earned higher scores in an essay placement test than would have been expected from their standardized test results, concluding that women write better than men, one possible reason for their (relatively) superior grade-point averages.

Recently, Bridgeman (1988) found sex differences in the predictive validity of the multiple-choice and essay questions of the Advanced Placement Examination in Biology for females but not for males. Using grades in undergraduate biology courses from 10 widely differing colleges, Bridgeman found that objective and essay scores were equally good predictors of undergraduate grades for males but that essays were significantly poorer predictors than objective items for females. Similar differences were found within the AP Examination itself, in that essay and objective sections correlated .59 for males but only .39 for females.

All these results may be related to differences in course selection by males and females, differences in instructors' grading of males and females, or some interaction of these. In any case, the difference in predictive validity bears further study.

Needless to say, results based on admission tests, whether tests for admission to undergraduate, graduate, or professional schools, do not reflect the general population. These results are based on self-selected samples that tend to be more able than the general population. Three sources of data from more nationally representative groups are the National Assessment of Educational Progress (NAEP); High School and Beyond (HS&B), a national longitudinal study of transition among high school students; and norming administrations of nationally normed standardized tests.

Tests Involving Nationally Representative Samples

The National Assessment of Educational Progress (NAEP) was initiated in 1969 to assess the level of achievement among students in the United States. In-school assessments are conducted on a periodic basis for 9-, 13-, and 17-year-old students. The subject areas assessed are rotated so that each assessment includes a different set of areas. For example, reading achievement has been assessed four times since 1969: in 1970–71, in 1974–75, in 1979–80, and in 1983–84. Mathematics was first assessed in 1972–73, then in 1977–78 and 1981–82, and most recently in 1985–86. Within each subject area, a variety of tasks measure performance on sets of objectives developed by panels of specialists.

The assessments are based on a deeply stratified, three-stage sampling design that produces large samples with approximately equal proportions of males and females at each age. Use of item response theory (IRT) in the past several assessments to estimate levels of proficiency has also provided a common scale on which to compare performance across time for the three age levels, as well as for subgroups within the population.

NAEP results from the 1986 assessment of mathematics showed little difference between the percentages of males and females at lower levels of proficiency among all three age groups. However, consistent with findings that show quantitative differences appearing around the time of adolescence, larger proportions of 13- and 17-year-old males achieved at the higher levels of proficiency. With respect to reading proficiency, females outperformed males at all three age levels in the 1984 assessment of that domain. This was true for all levels of reading proficiency, which are associated with increasing levels of reading complexity. However, results from an assessment of the literacy levels of adults aged 21–25 conducted in 1985 showed virtually no differences in the proficiency levels of men and women on any of the three scales used in the assessment (Mullis 1987).

Females also performed better than males on an assessment of writing conducted in 1984 among nearly 55,000 students in grades 4, 8, and 11. Fifteen

tasks were administered to students at each grade level and read by experienced teachers of English using specific guidelines for the evaluation of each task. On the basis of a measurement technique (Average Response Method, or ARM), that summarizes writing achievement on a common scale across grade levels, the superiority of female performance on the writing assessment appears to have increased in the higher grade levels (see Table 1).

The National Longitudinal Study (NLS) and High School and Beyond (HS&B) are longitudinal studies initiated by the National Center for Education Statistics. Based on nationally representative samples of high school sophomores and seniors, the studies collect information on the achievement, attitudes, aspirations, and future plans of the students, as well as demographic and family background data to provide context for the former. Test batteries administered in 1972 and 1980 included measures of vocabulary, associative memory based on picture-number pairs, reading comprehension, inductive reasoning based on letter groups (1972 only), mathematics, mosaic comparisons, and spatial relations in the form of three-dimensional visualization (1980 only). Each of the batteries, administered to randomly selected samples of students in over 1,000 randomly selected high schools, took over an hour. Factor analyses of the items yielded four major groups of items for which scores were reported: vocabulary, reading, mathematics, and science (Rock et al. 1985). In both years, the samples included slightly more females than males (50.1 percent compared with 49.9 percent in 1972 and 51.9 percent compared with 48.1 percent in 1980).

Item Response Theory (IRT) scaled scores for males and females in each of the cohorts for the vocabulary, reading, and math items appear in Table 2. The scores in reading and vocabulary are virtually identical for males and females, but math scores diverge for the two groups. In all cases, males score higher, but the magnitude of the difference is greater for the NLS seniors and the HS&B sophomores; for both the 1980 sophomores and the 1982 seniors, the differences are quite small.

Table 1. Average Writing Achievement (ARM)[a] Nation, Males, and Females

	Grade 4	*Grade 8*	*Grade 11*
Nation	158(1)[b]	205(1)	219(1)
Male	150(1)	196(1)	209(1)
Female	166(1)	214(1)	229(1)
Difference[c]	16	18	20

a. ARM, or Average Response Method, summarizes writing achievement on a common scale across grade levels. The scale ranges from 0 to 400.
b. Jackknifed standard errors are presented in parentheses.
c. Positive values for differences indicate that females had higher writing achievement.

Source: Mullis 1987

Table 2. NLS/HS&B IRT Mean Scaled Scores (High School Seniors)

	1972	*1982*
Vocabulary		
Total	6.55	5.76
Male	6.44	5.78
Female	6.67	5.75
Reading		
Total	9.89	8.13
Male	9.83	8.23
Female	9.95	8.03
Mathematics		
Total	12.94	11.43
Male	13.97	11.76
Female	12.09	11.09

Source: Rock et al. (1985)

In 1980 the Armed Services Vocational Aptitude Battery (ASVAB) was administered to a nationally representative sample of nearly 12,000 young people between the ages of 16 and 23, in order to develop national norms for the test. This sample was the same one that had been identified for inclusion in the National Longitudinal Study (NLS) of 1980. ASVAB is used "to determine eligibility for enlistment and qualification for assignment to specific military jobs" (Department of Defense 1982). The battery consists of 10 subtests: Arithmetic Reasoning, Numerical Operations, Paragraph Comprehension, Word Knowledge, Coding Speed, General Science, Mathematics Knowledge, Electronics Information, Mechanical Comprehension, and Automotive-Shop Information. Scores from four of these subtests—Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Numerical Operations—are combined to produce what is known as an Armed Forces Qualification Test (AFQT) score, which forms the basis for a decision about enlistment eligibility. Various subtests are also combined to form composites that determine eligibility for specific military fields. For example, an Administrative composite is made up of the Paragraph Comprehension, Word Knowledge, Numerical Operations, and Coding Speed subtests; and an Electronics composite is made up of the Electronics Information, General Science, Arithmetic Reasoning, and Mathematics Knowledge subtests.

Analysis of the 1980 data showed the percentile scores of males and females for the AFQT to be similar but average scores on various of the aptitude composites to differ. The AFQT claims to measure

verbal and quantitative abilities in equal proportion (Department of Defense 1982, p. 31); the overall mean AFQT percentile score for males was 50.8, for females 49.5. There were slight variations in these small differences across the three age groups distinguished by the study: females scored (insignificantly) higher at ages 18 and 19; males scored insignificantly higher at ages 20 and 21, but by ages 22 and 23, males surpassed females by four percentile points.

Males scored higher than females on the Mechanical, General, and Electronics composites; females scored higher than males on the Administrative composite. The widest gap between the scores of males and females involved the Mechanical composite, where the mean percentile score for males (51) was almost twice the mean percentile score for females (26). Males' score on the Electronics composite was 53 compared with females' 41, and on the General composite, 52 compared with 48. Females scored 51 on the Administrative composite, compared with males' 44. The mean estimated reading level, expressed as a grade equivalent, for the total sample of males was higher (9.6) than the score for females (9.3) by three months.

The Differential Aptitude Tests (DAT) are a battery of eight tests developed to measure the abilities of students in grades 8 through 12. There are measures labeled Verbal Reasoning, Numerical Ability, Abstract Reasoning, Clerical Speed and Accuracy, Mechanical Reasoning, Space Relations, Spelling, and Language Usage. Norms were developed for these tests using a representative sample of more than 60,000 public- and parochial-school students. Lupkowski (1987) computed standardized effect sizes for the mean difference by sex and ratios of males to females who scored at or above the ninetieth percentile on each of the eight tests, using data reported in the *Administrator's Handbook* (1982). Using Cohen's (1977) criterion, she found effect sizes considered "large" (d=.88 standard deviation), which favored twelfth-grade males on the Mechanical Reasoning test, and effect sizes considered "small" (d=.22) which also favored twelfth-grade males on the Space Relations test. On the Mechanical Reasoning test, effect sizes favoring males increased regularly between grades 8 and 12, from .66 at grade 8 through .88 at grade 12. The ratio of twelfth-grade males to females scoring at the ninetieth percentile on the Mechanical Reasoning Test was almost 5 to 1; for the Space Relations test, the ratio was almost 2 to 1. Lupkowski also found effect sizes in the "small to medium" range that consistently favored females in tests of Spelling (d ranging from .38 to .50), Language Usage (d ranging from .37 to .42), and Clerical Speed and Accuracy (d ranging from .29 to .37). The ratio of twelfth-grade males to females scoring at or above the ninetieth percentile on the Spelling test was 1 to 2.4. Effect sizes so small as to be considered negligible were found for tests of Numerical Ability, Abstract Reasoning, and Verbal Reasoning.

Verbal Ability

Prior to the Maccoby and Jacklin (1974) review, Anastasi (1958) and Tyler (1965), in widely read texts on differential psychology, claimed superiority of females over males in verbal functioning throughout the life cycle. Maccoby (1966) disagreed, claiming different relative advantages for females at different ages. For instance, she maintained that girls exceed boys through the preschool and early school years (speaking earlier, in longer sentences, and more fluently; learning to read sooner; and requiring less remediation in the process of learning to read), but that boys catch up, at least in reading, by about age 10. Further, she contended, girls tend to outperform boys on tests of grammar, spelling, and word fluency. Later, Maccoby and Jacklin (1974) claimed that more recent studies had shown few or no sex differences during the early years, but claimed evidence for a divergence between the sexes starting around age 11. Females scored higher on tasks involving receptive and productive language, fluency, analogies, comprehension of written material, and creative writing. This superiority of females was thought to increase through high school and possibly beyond, and, although the extent of the female advantage tended to vary with the study and the ability under scrutiny, the most commonly cited magnitude was about one-fourth of a standard deviation (Maccoby and Jacklin 1974, p. 351). Other reviews (Denny 1982; Halpern 1986) concurred with these conclusions.

Although these reviews agree that there *are* gender differences in verbal ability, they disagree about the kinds of verbal tasks that show such differences and also about the nature of developmental trends in gender differences. It would appear that the size of the difference and the direction of the difference vary with individual studies and depend on the age of the test-takers, the ability or abilities tested, the sample of test-takers represented by the data, and even the decade in which the study was conducted. In fact, many of the studies reviewed by Maccoby and Jacklin showed no differences at all in the performances of males and females, and many showed only small differences.

NAEP data, for instance, showed superiority of women tested in school in both reading and writing performance. However, when a literacy assessment was conducted on young adults ages 21 to 25, the gender difference had disappeared. About 3,600 individuals were interviewed at home as part of the literacy assessment. They were asked to respond to 100 tasks organized into three scales, one of which contained tasks from the NAEP reading assessment.

Table 3. National Assessment of Educational Progress Literacy Levels for Young Adults (21–25 Years Old), 1985[a]

	Male	*Female*	*Difference*
NAEP reading proficiency	304.6(2.3)	305.4(2.3)	0.8
Prose comprehension	305.6(2.6)	304.5(2.1)	−1.1
Documents	305.3(2.6)	304.8(1.9)	−0.5

a. Jackknifed standard errors are presented in parentheses. Positive values for differences indicate that females had higher literacy levels.

Note: Sampling methods (exclusion of prisons and military bases) may have affected the results slightly more for males than for females.

Source: Mullis 1987

The IRT results for the three literacy scales (for which the range is from 200 to 500) show virtually no difference between men and women with respect to their proficiency levels on any of the three scales (see Table 3).

For both men and women, the level of performance on the NAEP reading scale was higher than the level attained by in-school 17-year-olds in 1984. The nine-point difference in performance between males and females in the latter group did not show up in the results for young adults tested in their homes.

Meta-Analysis of Verbal Differences

Using as a starting point the assertion that little is really known about the nature of gender differences in verbal ability, Hyde and Linn (1988) undertook a meta-analysis of existing primary research reports. Their analysis included both the studies that Maccoby and Jacklin had used and a group conducted after that review, a total of 165 different studies. Their goals, in addition to assessing the magnitude of gender differences generally, included assessing differences on different measures of verbal ability, trends in these differences, and possible differences in cognitive processing that might account for any observed differences.

Hyde and Linn report a small positive weighted mean value (signaling superior performance by females) of the difference (d) in male–female performance averaged over 119 available values.[2] More of the studies (75 percent) reflected superior female performance, although only 27 percent of them attributed statistical significance to this finding; 7 percent of the studies found males performing better than females at a statistically significant level. More important, the set of effect sizes was judged (against a statistical criterion) to be heterogeneous, suggesting that different kinds of ability were being assessed.

Grouping the studies by type of ability tested (vocabulary, analogies, reading comprehension, verbal communication, essay writing, general verbal ability—except for the SAT-verbal sections, which became its own category—anagrams, and "other"), Hyde and Linn found the magnitude of gender differences to be close to zero for many types of tests. The only exceptions were modest (about .20 standard deviation) differences favoring females on measures of general verbal ability and the solution of anagrams, and a difference of about .33 standard deviation favoring females in measures of the quality of speech production. A cognitive processing analysis identified five processes that might be involved in the various tests of verbal ability: retrieval of a word definition, retrieval of the name of a picture, analysis of the relationships among words, selection of relevant information from an information source such as a reading passage, and verbal production; some combination of these; and "other." As was the case for types of tests, the effect sizes for different cognitive processes were small (the largest was $d = +.19$, for both combinations of processes and "other"); five of the seven favored females. With the possible exception of some slight superiority of males aged 6 to 10 on tests of vocabulary, all the effect sizes in an analysis by type of test and age were essentially uniform across age groupings, and small and positive (that is, favoring females). Linn and Hyde conclude,

> We are prepared to assert that there are no gender differences in verbal ability, at least at this time, in this culture, in the standard ways that verbal ability has been measured. We feel that we can reach this conclusion with some confidence, having surveyed 165 studies, which represent the testing of ... (excluding the ... SAT data ...) 441,538 subjects and averaged 119 values of d to obtain a weighted mean value of +0.11. A gender difference of one-tenth of a standard deviation is scarcely one that deserves continued attention in theory, research or textbooks. Surely we have larger effects to pursue. (Hyde and Linn 1988, p. 23)

Quantitative Abilities

Like verbal ability, quantitative ability is a general label that incorporates a number of different areas of competence. For example, the "quantitative" sections of measures of "long-range achievement" or "ability," like the SAT and the GRE, in which men consistently outperform women, refer almost exclusively to mathematics. Other measures, like the DAT, include sections that tap spatial ability. The

2. Actually, they report a small positive *un*weighted mean value but a very small negative value (signaling slightly superior male performance) when effect size was weighted by number of subjects. This shift was attributable to a single study with a negative value of d, based on just under one million SAT-takers (Ramist and Arbeiter 1986). The study was removed from further consideration in the meta-analyses and treated and discussed separately in the results.

latter is often hypothesized to contribute to the former. That is, gender differences in spatial ability are often invoked to explain gender differences in mathematics. Benbow and Stanley (1980, p. 1263), for example, claim that "sex differences in achievement in and attitude toward mathematics result from superior male mathematical ability which may, in turn, be related to greater male ability in spatial tasks."

Mathematics Ability and/or Achievement. Studies of mathematical ability and achievement have consistently found sex differences favoring males among high school students (Fennema 1974; Benbow 1988; and others) and claim that the differences first become apparent in junior high school. Girls are generally believed to excel in computation, boys in tasks that require mathematical reasoning.

As noted earlier, the 1986 NAEP assessment of mathematics achievement found few differences in the performance of men and women at the lower levels of proficiency but larger proportions of 13- and 17-year-old males achieving at higher levels of proficiency. There was a consistent advantage for males on the geometry scale at grades 7 and 11 and on the measurement scale at all three grade levels, statistically significant at grades 3 and 11, but not at grade 7. At all three grade levels, there was a consistent advantage for females in the areas of knowledge and skills and a consistent advantage for males in the area of "higher-level applications." Females tended to outperform males on tasks "where there is an obvious procedural rule to follow," but males had an advantage when the strategy for solving the problem was less apparent. There were no gender differences with respect to the algebra subscale (Dossey et al. 1988).

Along with achievement data, NAEP collects information about students' attitudes toward and perceptions of mathematics. Males and females in grade 3 responded to two questions about their enjoyment of mathematics (60 percent agreed with the statement "I like mathematics," and 40 percent of the males and 43 percent of the females responded positively to the statement "I would like to work at a job using mathematics") and their confidence in their mathematical abilities (65 percent of the males and 66 percent of the females agreed with the statement "I am good with numbers"). Within an overall pattern of decreasing interest with age in mathematics and confidence in their own abilities, females became slightly more negative than males on both counts in grades 7 and 11.

Benbow and Stanley (1980, 1983b) have been studying the performance and related characteristics of samples of intellectually talented students in their Study of Mathematically Precocious Youth (SMPY) since 1972. Although the SMPY program was broadened in 1980 to include verbally as well at mathematically precocious students, the selection procedures have remained essentially the same. The students, the majority of whom are seventh and eighth graders, are identified by means of their performance on various standardized tests administered in schools, as meeting some minimum criterion (ninety-fifth, ninety-seventh, or ninety-eighth percentile). These high-scoring students are then invited to take the SAT in order to qualify for a range of special programs. Since not all schools nominate students and not all students identified by their test scores agree to take the SAT, the SMPY population is a self-selected one. A survey of one cohort of such students (Wilder, Casserly, and Burton 1988) revealed them to be largely white and of middle-class origin. Nonetheless, Benbow and Stanley have accumulated data representing more than 100,000 SMPY students since the start of the program in 1972, data that reflect test performance and a large number of background and attitude variables.

Throughout that period, the SMPY students showed no significant mean gender differences in verbal ability, but consistent differences on the SAT-mathematical sections on the order of 30 points, favoring males. For example, in a report based on the testing of about 50,000 students between 1972 and 1982, consistent differences of at least 30 points in mean SAT-mathematical scores were found for males and females, and far more boys than girls achieved high scores on the test.

Spatial Ability. Studies of gender differences in mathematical performance often make distinctions between items that do and do not involve spatial ability, claiming greater superiority for males over females in the former (see, e.g., Fennema and Carpenter 1981). As part of their study of mathematically precocious youth, Benbow at al. (1983) tested some of the most precocious of their sample with a special battery of specific mental-ability measures. They identified two factors that were able to account for these students' extraordinary performance, one a verbal factor and the other spatial, a finding which, they claimed, "implicates the importance of spatial ability in accounting for the high level test performance of these mathematically talented students" (Benbow 1988, p. 23).

Spatial ability itself is defined and studied in a variety of ways. In their meta-analysis of spatial differences reported after Maccoby and Jacklin's 1974 review (and before 1982), Linn and Peterson (1986) identified four different perspectives that distinguish research on spatial ability: the differential, concerned with performance differences among different populations; the psychometric, concerned with identifying the "structure" of the spatial domain; the

cognitive, concerned with identifying the processes used to solve spatial tasks; and the strategic, concerned with identifying strategies used by test-takers attempting to solve spatial tasks.

After reviewing studies across perspectives, Linn and Peterson divided the spatial domain into three broad categories which they labeled "spatial perception," "mental rotation," and "spatial visualization" (Linn and Peterson 1986, p. 70). Spatial perception tasks require subjects to locate true horizontal or vertical in the presence of distracting information. Examples include the Rod and Frame Test (RFT), which asks the test-taker to situate a rod in a vertical position in the presence of a frame oriented at an angle; and water-level tasks that ask the test-taker to draw or identify a horizontal line in a tilted bottle. Mental rotation involves the ability to rotate (in the mind) a two- or three-dimensional figure. Spatial visualization refers to tasks that require analytic processing of spatially presented information, for example, locating embedded figures, block design, and paper folding.

Linn and Peterson computed and tested a total of 172 effect sizes and, finding a lack of homogeneity, partitioned them into not only the three categories described above but also three age groups (under 12, 12 to 17, and 18 or older) within each category. For tasks involving spatial perception, they found differences favoring males among individuals as young as 7 or 8. These differences increased with age, reflected in weighted effect sizes of .37 standard deviation for the under 12 and 12 to 17 groups, and of .64 for those over 18. Likewise, gender difference on mental rotation tasks were found throughout the life span, although, because of difficulties involved in testing younger subjects for mental rotation, the domain has not been measured with children younger than 10. Gender differences favoring males in spatial visualization were found to be so small (on the order of .13 of a standard deviation) as to be considered neither significant nor meaningful, and consistently so across the three age groups.

Thus there is evidence of differences in the test performance of males and females on some tasks on some tests. These differences are quite small in the verbal domain and larger in quantitative areas. The quantitative differences seem to appear or increase during the high school years and may involve differences in spatial perception and/or mental rotation.

TRENDS IN SEX DIFFERENCES IN PERFORMANCE

Many of the studies cited, and almost all of the studies included in meta-analyses, reflect the performance of single cohorts of test-takers at one point in time. It is also instructive to examine changes in test scores over time. It is possible, for example, that changes in the educational and social experiences of women engendered by increased opportunities for coeducation and by the women's movement have affected women's test performance and their standing relative to men. Many of the tests described in earlier sections of this report have equated forms that are administered periodically. Advances in item response theory (IRT) have made it possible to consider the results from different years of national testing programs, like NAEP, on a common scale. Finally, comparing the results over time for several different testing programs can illuminate performance trends across different populations.

Voluntary Testing Programs

Burton (1987) examined trend data over two decades for several voluntary testing programs and found that women's scores on the verbal components had declined in all of them relative to men's. On some tests, specifically the Test of Standard Written English (TSWE), the English Composition Test (ECT), the American College Testing Program (ACT) English Test, and the verbal portion of the Graduate Management Admissions Test (GMAT-V), women continued to earn higher scores than men through the 1980s, but the score difference between the sexes declined. In other programs, specifically the vocabulary and reading comprehension sections of the Scholastic Aptitude Test, the verbal section of the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT), ACT social studies and natural science tests, and the verbal portion of the Graduate Record Examination (GRE-V), women were scoring lower than men by the mid- to late 1970s and the score difference between men and women continued to widen through the mid-1980s.

Burton also reported a general increase in the relative number of women participating in all of the voluntary testing programs she examined, even in those (ECT and GMAT) that suffered losses in total volume over some or all of the years included in her study. Overall, she found that, at least during the 1970s, as the proportion of women in each testing program grew, their scores relative to men's declined. This relationship was less marked for the undergraduate admission tests during the 1980s, although it did continue for the graduate-level tests. And the data for each testing program showed different slopes (rates of change) and intercepts (sex differences at the start of the period of interest). Burton (1987, p. 6) concluded that the differences were probably due "both to the differences in ability and demographic profiles among testing populations, the differences

in the constructs being measured, and the differences in difficulty of the various tests."

Data from the Law School Admissions Council/ Law School Admissions Service (LSAC/LSAS 1988) for the six-year period between 1981-82 and 1986-87 showed an almost negligible decline, from 31.6 to 31.4 for males and from 31.3 to 31.2 for females, along with declines in average GPA and a drop in the number of applicants through 1985-86 (see Table 4). Relatively small declines were also observed in MCAT Reading and Quantitative scores for both males and females between 1978 and 1983 (see Table 5). Over the same period, scores on the various science portions of the test *increased* for both males and females, although females' scores remained lower than males' throughout, for all the science tests—biology, chemistry, physics, and science problems (Jones 1984). A more recent set of single-year results (AAMC 1987) showed a recent drop in scores across all MCAT tests and subgroups.[3] Females retained their relative disadvantage in all but the reading subtest, but there were reductions in the magnitude of the male-female difference.

Table 4. Mean LSAT Scores and GPA, by Sex

	LSAT[a]				*GPA*[b]	
	Males		*Females*		*Males*	*Females*
	n	*Score*	*n*	*Score*	*Score*	*Score*
1981-82	44,981	31.6	27,918	31.3	3.05	3.17
1982-83	44,051	31.6	27,783	31.5	3.02	3.15
1983-84	38,482	31.7	25,284	31.4	3.03	3.15
1984-85	35,987	31.6	24,325	31.1	3.02	3.13
1985-86	38,349	31.6	26,606	31.2	3.00	3.11
1986-87	39,729	31.4	28,765	31.2	2.99	3.10

a. The LSAT scale ranges from 13 to 49.
b. GPA is computed by the LSAS from student transcripts.
Source: LSAC/LSAS National Statistical Report (1988)

Table 5. Mean MCAT Scores of Male and Female Medical School Applicants

	Applicant Year						
MCAT	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*...1987*
Biology							
Males	8.6	8.7	8.6	8.8	9.0	9.2	8.3
Females	8.0	8.3	8.2	8.3	8.4	8.6	7.8
Chemistry							
Males	8.7	8.7	8.9	8.8	8.9	9.0	8.1
Females	7.7	7.8	8.0	8.0	8.0	8.1	7.4
Physics							
Males	8.8	9.0	8.7	9.0	9.2	9.2	8.3
Females	7.4	7.6	7.4	7.6	7.8	8.0	7.2
Science Problems							
Males	8.8	8.9	8.8	8.8	8.9	9.0	8.1
Females	7.7	8.0	7.9	7.8	7.9	8.0	7.3
SA: Reading							
Males	8.4	8.5	8.3	8.2	8.1	8.3	7.5
Females	8.5	8.6	8.4	8.3	8.1	8.2	7.5
SA: Quantitative							
Males	8.7	8.7	8.5	8.4	8.3	8.3	7.8
Females	7.8	8.0	7.8	7.6	7.5	7.4	7.1

Source: Division of Student Services, Association of American Medical Colleges.

Nationally Representative Samples

It is once again instructive to examine the data from tests administered to nationally representative samples of students. Such data include the results from the National Assessment of Educational Progress (NAEP) and the National Longitudinal Study of the high school graduating class of 1972 (NLS) and its successor, High School and Beyond (HS&B), which do not suffer from the limitations of self-selected samples; and data from the standardized administrations of nationally normed tests (e.g., the Differential Aptitude Tests (DAT).

National Assessment of Educational Progress

Reading. In the national assessment of reading in 1971, there was a 12-point difference between the performance levels of males and females at all three age levels (Mullis 1987). In 1984 the difference was smaller at the three age levels, but in differing degrees and at different times for each age level (see Table 6).

A comparison across the four assessments between 1971 and 1984 reveals that the reading proficiency of males has trailed that of females in all four assessments for all three age groups, but that the gap between the two groups narrowed slightly over the 13-year period. The proficiency of 9-year-olds improved significantly during the period between 1971 and 1980 but showed no improvement between 1980 and 1984; in fact the proficiency of females in this group dropped between 1980 and 1984, so that the 12-point difference between the two groups in 1971 had been reduced by nearly half to a six-point difference in 1984 (Mullis 1987).

A similar, but not quite so dramatic, narrowing of the gap also occurred among 13-year-olds. Again, males improved more than females did during this period.

[3]This shift suggests that data for the years between 1983 and 1987 be examined for information about the nature of the shift in direction.

Table 6. National Assessment of Educational Progress Trends in Mean Reading Proficiency for the Nation, Males, and Females

	1971	*1975*	*1980*	*1984*
Age 9				
Nation	207.2(1.1)[a,b]	209.6(0.7)[a]	213.5(1.1)	213.2(0.9)
Male	201.2(1.2)[a]	204.2(0.9)[a]	208.5(1.2)	210.0(1.0)
Female	213.3(1.2)	215.1(0.8)	218.5(1.1)	216.3(0.9)
Difference[c]	12.1	10.9	10.0	6.3
Age 13				
Nation	253.9(1.1)[a]	254.8(0.8)[a]	257.4(0.9)	257.8(0.6)
Male	247.9(1.1)[a]	248.4(0.8)[a]	252.8(1.1)	253.5(0.7)
Female	259.9(1.1)	261.2(0.9)	261.8(0.9)	262.3(0.7)
Difference	12.0	12.8	9.0	8.8
Age 17				
Nation	284.3(1.2)[a]	284.5(0.7)[a]	284.5(1.1)[a]	288.2(0.9)
Male	278.1(1.2)[a]	279.2(0.8)[a]	281.1(1.2)	283.4(0.9)
Female	290.3(1.3)	289.6(0.8)[a]	287.9(1.2)[a]	293.1(1.0)
Difference	12.2	10.4	6.8	9.7

a. Significantly different from 1984.
b. Jackknifed standard errors are presented in parentheses.
c. Positive values for differences indicate that females had higher proficiency levels.
Source: Mullis 1987

Trends in reading achievement for 17-year-olds differed quite strikingly from the trends for the other two groups (reflecting, at least in part, the phenomenon of drop-out, since NAEP assessments are administered to in-school populations). The level of reading proficiency for 17-year-olds remained quite constant throughout the 1970s but showed a significant improvement between 1980 and 1984. Males showed steady improvement over the 13-year interval; females, on the other hand, showed declines in performance throughout the 1970s but a dramatic improvement between 1980 and 1984. Thus the discrepancy between 17-year-old males and females, which was smallest (at 6.8 points) in 1980, increased in 1984, although not to the magnitude of 1971. Expressed in terms of reading levels, the percentage of 17-year-old males reading at the "adept" level of proficiency (the fourth of five scaled levels) was 32 percent between 1971 and 1980, and rose to 35 percent in 1984. The proportion of females at this level declined from 43 in 1971 to 38 in 1980, then returned to 44 in 1984 (NAEP 1985).

With respect to writing, women have had consistently higher scores than men across the three different writing tasks for which comparisons could be made between results obtained in the 1979 and 1984 assessments (see Table 7). This table shows the differences between males and females in the percentages scoring at the levels given (2, 3, or 4 in the primary trait analysis and 4, 5, or 6 in the holistic scoring). Apart from the consistently superior performance of women, there does not appear to be a clear trend in the data. Differences appear to have diminished among 17-year-olds but vary according to task and level (primary trait or holistic) for the other two ages.

Mathematics. Mathematics achievement has been assessed four times since the inception of NAEP: in 1973, 1978, 1982, and 1986. Using IRT scaling models and a common scale with five anchor points, and extrapolating to the assessment of 1973 in which fewer common items were used than in the other three, it is possible to examine trends over a 15-year period. Overall, these trends show some decline in mathematics achievement between 1973 and 1978, and modest gains thereafter (Dossey et al. 1988). Trends in average proficiency levels look similar for males and females, although there were some subtle shifts. For instance, although average proficiency levels in 1986 were virtually identical for 9-year-old boys and girls and for 13-year-old boys and girls, these levels represent significant gains since 1978 for boys but not for girls (see Table 8). That is, girls' performance remained comparatively consistent across the years, whereas boys' improved. Among 17-year-olds, the situation was roughly the opposite. In each of the four assessments, the mathematics achievement of males was higher than that of females, but, in recent years particularly, females may have begun to close the gap. The performance of all 17-year-olds, male and female, declined between 1973 and 1982 but had improved by 1986. That improvement was statistically significant for females but not

Table 7. National Assessment of Educational Progress Trends in Male/Female Differences on Three Writing Tasks

	Primary trait (2,3,4)		*Holistic scoring (4,5,6)*	
	Difference in 1979	*Difference in 1984*	*Difference in 1979*	*Difference in 1984*
Age 9				
Informative	12.3[a]	9.8	12.2	17.2
Persuasive	9.9	14.7	15.1	23.7
Imaginative	9.3	7.5	18.8	13.4
Age 13				
Informative	12.1	7.6	0.3	1.3
Persuasive	0.1	1.0	3.8	7.9
Imaginative	10.1	12.0	23.7	15.3
Age 17				
Informative	13.3	10.6	23.4	16.3
Persuasive	3.7	3.1	20.7	14.9
Imaginative	10.7	8.9	22.1	19.7

a. Positive values for differences indicate that females have higher writing achievement.

Source: Mullis 1987

Table 8. Trends in Average Mathematics Proficiency for 9-, 13-, and 17-Year-Olds by Gender

	Females			*Males*		
	Age 9	*Age 13*	*Age 17*	*Age 9*	*Age 13*	*Age 17*
1973	[220.4][a]	[266.9]	[300.6]	[217.7]	[265.1]	[308.5]
1978	219.9(1.0)[b]	264.7(1.1)	297.1(1.0)	217.4(0.7)[c]	263.6(1.3)[c]	303.8(1.0)
1982	220.8(1.2)	268.0(1.1)	295.6(1.0)	217.1(1.2)	269.2(1.4)	301.5(1.0)
1986	221.7(1.2)	268.0(1.5)	299.4(1.0)	221.7(1.1)	270.0(1.1)	304.7(1.2)

a. Brackets indicate that data were extrapolated from previous NAEP analyses.

b. Jackknifed standard errors are presented in parentheses.

c. Statistically significant difference from 1986 at the .05 level.

Source: Dossey et al. 1988

for males, so that by 1986 there were negligible gender differences in mathematics proficiency at all three levels, but the small differences were relatively larger for 17-year-olds than for other groups. At the same time, trend data reflecting the Assessments of 1978, 1982, and 1986 show females at ages 13 and 17 expressing increasing confidence in their mathematical abilities.

National Longitudinal Study (NLS) and High School and Beyond (HS&B)

Ekstrom et al. (1988) examined trends in the nationally representative data collected for the NLS Study of 1972 and its successor, HS&B. These data provide cross-sectional estimates of the achievement of students who were high school seniors in 1972 and in 1982, in three areas: vocabulary, reading, and mathematics. About half of the items were identical in 1972 and 1982; item response theory (IRT) equating was used to put the 1972 and 1982 test scores on a common scale to facilitate comparison between the two. The data are summarized in Table 9. The table shows that scores declined over the 10-year period, for the total group and for males and females (and, in fact, for almost every subgroup examined in the Ekstrom et al. analysis). In all cases low-socioeducational status students showed larger score declines than high SES students, students in public schools showed larger declines than students in nonpublic schools, and students in the general and vocational curriculums showed greater declines than students in the academic curriculum (Ekstrom et al. 1988, p. 75). Against this general background of gloom, the apparent convergence of the scores of males and females seems like good news. With respect to vocabulary and reading, males' scores declined less than females,' to the point where females lost their (only slight) advantage. In mathematics, males' scores declined considerably more than females,' although males continued to score higher than females in 1982.

Table 9. Changes in NLS/HS&B Mean Test Scores (High School Seniors)

	1972	*1982*	*Difference*	*Effect Size*
Vocabulary				
Total	6.55	5.76	-0.79[a]	-0.20
Male	6.44	5.78	-0.66[a]	-0.17
Female	6.67	5.75	-0.92[a]	-0.23
Reading				
Total	9.89	8.13	-1.76[a]	-0.34
Male	9.83	8.23	-1.60[a]	-0.31
Female	9.95	8.03	-1.92[a]	-0.38
Mathematics				
Total	12.94	11.43	-1.51[a]	-0.20
Male	13.97	11.76	-2.30[a]	-0.27
Female	12.09	11.09	-1.00[a]	-0.14

a. Statistically significant difference.
Source: Ekstrom et al. 1988.

Ekstrom el al. also compared the changes in tested achievement in standard deviation units for the NLS72 and HS&B populations with the population of students that took the SAT in 1972 and 1982. The comparison appears in Table 10. In both groups, the decline is greater in the verbal/reading area than in mathematics. Overall, except for the math scores of females, the decline was greater for the NLS-HS&B population than for the SAT-takers.

The Differential Aptitude Test and PSAT/NMSQT

Contending that cognitive gender differences are disappearing, Feingold (1988) examined the norms from four standardizations of the DAT conducted between 1947 and 1980 and from four standardizations of the PSAT/NMSQT and SAT conducted between 1960 and 1983. Gender differences had been found on all DAT and PSAT/NMSQT subtests when the instruments were normed in 1947 and 1960, respectively, females generally scoring higher on verbal measures, males on quantitative. The DAT scales, it will be recalled, assess verbal reasoning, spelling, and language in the verbal domain; and clerical speed and accuracy (perceptual speed), space relations (three-dimensional spatial visualization), numerical ability (arithmetic), mechanical aptitude, and abstract reasoning in the quantitative domain. The PSAT/NMSQT, like the SAT to which it is related, provides two scores, verbal and quantitative. Presumably, the populations chosen for standardization of the DAT were nationally representative. The four PSAT/NMSQT populations were also representative samples of high school juniors and seniors. By way of contrast, average PSAT/NMSQT and SAT scores derived from yearly program data describe self-selected populations of college-bound high school juniors and seniors. The use of these three populations allows for comparisons between the group of self-selected students whose scores comprise the averages quoted for undergraduate admission examinations, and the more representative groups of students who comprise the standardization populations for nationally normed tests.

Feingold's analytic procedures involved standardizing gender differences over grade and year of examination to obtain mean effect sizes for each ability. Analysis of data from the DAT showed consistent superiority of females over males in tests of spelling, language, and clerical speed and accuracy. In spelling, females' advantage increased steadily from grade 8 to grade 12. In all three tests, however, females' relative advantage declined significantly over the period in question, from 1947 to 1980. Males scored higher than females on tests of mechanical reason-

Table 10. Changes in SAT and NLS-HS&B Test Scores, 1972–1982

	SAT test score changes, 1972–1982							
	Verbal				*Mathematical*			
	1972	*1982*	*Diff.*	*Change in SD units*	*1972*	*1982*	*Diff.*	*Change in SD units*
Male	454	431	-.23	-.21	505	493	-12	-.10
Female	452	421	-.31	-.28	461	443	-18	-.16
	NLS-HS&B test score changes, 1972–1982							
	Reading				*Mathematics*			
	1972	*1982*	*Diff.*	*Change in SD units*	*1972*	*1982*	*Diff.*	*Change in SD units*
Male	9.83	8.23	-1.60[a]	-.31	13.79	11.76	-2.30[a]	-.27
Female	9.95	8.03	-1.92[a]	-.38	12.09	11.09	-1.00[a]	-.14

a. Statistically significant difference.
Source: Ekstrom, et al. 1988.

ing and space relations for all grades and years, and gained more with respect to these abilities than girls did between grades 8 and 12. However, boys' relative advantage over girls diminished between 1947 and 1980 in terms of both scores and increases in relative performance. No appreciable gender differences were found for tests of verbal reasoning, abstract reasoning, and numerical ability.

With respect to the PSAT/NMSQT, although males consistently outperformed females on PSAT/NMSQT-mathematical sections, Feingold noted a decrease in effect size from .34 to .12 between 1960 and 1983 among juniors in the norming sample, the only group that was tested all four times. This change was due mainly to a decrease in the proportion of low-scoring females. However, although the gender difference had declined between 1960 and 1974 by 50 percent, on separate-sex norms, males needed to score about 50 points higher than females to achieve the ninety-ninth percentile in both years, suggesting that mathematically talented students remained disproportionately male. The small gender difference favoring females on PSAT/NMSQT-verbal sections in 1960 had virtually disappeared by 1974.

Feingold reports a difference of about one standard deviation in the verbal and mathematical scores of PSAT/NMSQT and SAT examinees, reflecting the self-selection bias inherent in SAT scores. Although gender differences for SAT-verbal scores were small among both juniors and seniors in all four years, Feingold notes the trends cited elsewhere of greater male gains relative to females on this portion of the test.

Are Gender Differences Disappearing?

Feingold's (1988) conclusion from examinations of the DAT and PSAT/NMSQT data is that females may indeed have narrowed the gender gap in quantitative performance over the past 20 or 40 years. At the same time, examination of both PSAT/NMSQT and SAT data over those periods shows that the math gender difference became more pronounced at higher levels of mathematical ability. For example, in 1981–82, 56 percent of the scores of 600, 81 percent of the scores between 750 and 770, 90 percent of the scores between 780 and 790, and 96 percent of the scores of 800 were earned by boys (Dorans and Livingston 1987). These findings are corroborated by Benbow and Stanley's data accumulated from 1972 (Benbow 1988). Feingold attributes this difference, at least in part, to the greater variability in male performance (expressed in consistently higher standard deviations of male examinees on both PSAT/NMSQT-mathematical and SAT-mathematical sections), which obliterated gender differences at the low end of the scale and magnified them at the high end. Benbow and Stanley have intensified their search for biological explanations for the phenomenon (Benbow 1988).

Trend data for the various testing programs reviewed here do seem to point to a narrowing or closing of the gap between males and females over the past 10 or 15 years. The historical small advantage enjoyed by females in the verbal domain appears to have been eliminated or, in some cases, reversed. The superiority of males in selected areas of the quantitative domain remains for some measures but appears to be less substantial than in the past. Even where the overall trend is downward, as it has been with NLS and HS&B data, males' and females' scores appear to be converging. The one area in which a strong male advantage remains is in the upper ranges of tested mathematics performance, exemplified by high scorers on the SAT-mathematical sections.

EFFORTS AT EXPLANATION

Efforts to explain both the differences in test scores and the trends in these differences run the gamut from the biological to the psychosocial, and from assertions of inherent, biologically based differences between males and females through critical assessments of differences in the social and educational experiences provided them to characteristics of the tests that show differences in performance. In recent years researchers have developed complex models to represent the interaction of many of the above variables in the genesis of differences in performance.

Many efforts to explain the differences treat them as "real" and seek to justify such treatment by identifying the mechanisms that underlie them. Other efforts regard the differences as "artifacts" of the differential treatment of men and women by our society—differences in socialization, experience at all levels of the educational process, aspirations, and expectations, to name some of the variables that have been hypothesized as responsible for the measured differences. Still other efforts attempt to explain the differences out of existence, by impugning the evidence, claiming lack of statistical or practical significance for it, or claiming bias in the measures that show differences. Efforts to explain trends in the data lean toward demographic variables and changes in patterns of social and educational phenomena. Finally, data from large-scale data-collection efforts like the National Longitudinal Study (NLS), High School and Beyond (HS&B), and other field studies have been used to construct and test models based on the interaction of many variables in the etiology of sex differences.

Biological Explanations

Although considerable media attention has been given to claims for biological bases of gender differences in cognitive abilities, and although there are large bodies of research on sexual dimorphism, hormonal influences, and other related topics in the biological bases of behavior, evidence for the relationship of these to cognitive abilities is still contradictory and incomplete. Biological explanations of sex differences center on three systems that might be responsible for cognitive differences: genetic or chromosomal determinants of sex-linked behaviors, differences in sex hormones secreted by the endocrine glands, and differences in the structure, organization, or function of the brain (Halpern 1986). Most of the studies and theories that address the biological bases of gender differences in cognitive abilities have focused on spatial abilities.

Although at least one study of spatial ability in twins (Vandenberg 1968) produced high correlations between pairs on measures of visualization and mental rotation, controversy remains over whether such correlations reflect heritability or merely the similar environmental conditions that pertain to twins; further questions exist about whether heritability explains sex differences in whatever phenomena are under investigation. Genetic explanations of sex differences in cognitive abilities would need to identify a mechanism of inheritance that is differentiated by sex (Halpern 1986, p. 70). A theory of sex-linked recessive genes accounting for differences in spatial ability gained some attention in the early 1960s, but its validity has since been seriously questioned. A parallel theory of a recessive trait linked to the X chromosome was proposed for verbal ability (Lehrke 1974) based on observed familial patterns of certain mental deficiencies. These observations do not appear to be equally applicable to individuals within the normal range of intelligence. Vandenberg (1968) also found high correlations in pairs of twins in verbal ability, concluding that verbal abilities, too, have a high heritability component. At the same time, he concluded that verbal abilities are more influenced by environmental factors than are spatial abilities, a conclusion that has been supported by more recent research. Stafford (1972) offered a model of heritability of mathematical abilities, based on patterns of intercorrelations among family members and a mechanism of a recessive X-linked gene. Sherman (1978) discredited both the data and the model.

There are no gross anatomical differences between male and female brains (Halpern 1986, p. 75); there are, however, some differences between the two. There are structural (and hormonal) differences related to the fact that women menstruate and men do not. In addition, women's brains are slightly smaller than men's (related, no doubt, to the fact that brain size among normal humans is correlated with body size). Neither of these differences has been reliably associated with differences in cognitive functioning.

Considerable research attention has focused on the question of differential hemispheric brain function and cognitive differences in the sexes. This research, as Halpern (1986) points out, derives from the observation that the functions for which the halves of the brain appear to be specialized (language, or verbal, ability in the left hemisphere and nonverbal, or spatial, ability in the right) are the very functions for which male–female differences have been documented. A review of the rather rich research in this area is beyond the scope of this paper. In its stead the reader is referred to reviews by Annett (1980), Hyde (1985), Halpern (1986), and Fausto-Sterling (1985), which, incidentally, differ in their conclusions. The views of Halpern and Fausto-Sterling are simplistically summarized here. Both reviewers agree that there is evidence of some differences in the organization and structure of the brains of males and females, of which handedness is one obvious manifestation. The existence of sex-by-handedness interactions in cognitive functioning suggests some role for neurological differences. At the same time, Halpern concludes, "The practical effect of sex differences in cerebral lateralization needs to be interpreted in a social context in which skills and abilities are prized and encouraged in a sex differentiated manner" (Halpern 1986, p. 90).

A major biological difference between men and women is in the relative concentrations of male hormones, mainly testosterone, and female hormones, estrogen and progesterone, that circulate throughout the bloodstream and affect behavior in several domains. All these hormones are present in both sexes, and the relative concentrations of all of them vary by sex and throughout the life cycle. Moreover, at least beginning in adolescence, the cyclical patterns of hormonal concentrations differ for men and women. Both the preponderance of a given hormone in one group or the other and the differences in cyclical patterns of hormone concentration have been made the basis for one or another causal theory of gender differences in cognitive abilities. In particular, changes in the relative concentrations of hormones and in the emergence of different hormonal cycles for males and females have been associated with changes in measured patterns of cognitive abilities (particularly mathematical and spatial abilities) that occur around adolescence. Such research suggests that some relationship may exist between sex hormones and cognitive abilities, but

the nature of the relationship has not been documented. Nor do existing theories deal satisfactorily with differences that exist prior to adolescence. And the hormonal changes that are implicated in such theories are confounded with significant life changes that occur simultaneously, seriously confounding any attempt to rule out environmental effects.

Nonetheless, Benbow (1988, in press), reviewing data from many years of her work with Julian Stanley on students they tested as part of the Study of Mathematically Precocious Youth (SMPY), maintains that the usual environmental explanations do not account adequately for the gender differences observed among more than 100,000 junior high school students who took the SAT as part of the SMPY process over a 15-year period. The SMPY population, it will be recalled, is a self-selected group of high-achieving 13-year-olds who take the SAT voluntarily, having been invited to do so by virtue of their high scores on standardized achievement tests. In the context of her most recent summary of these data, Benbow reviewed all the conventional environmental explanations for gender differences in test performance in math: negative attitudes of females toward math, females' lesser confidence in their own math abilities, math anxiety on the part of females, parents' and teachers' encouragement of males more than females in math, patterns of course-taking, and so on. She then showed, for each explanation, how SMPY students have not conformed to most of the expectations, or that the differences between males and females in the SMPY population have been smaller than differences observed in other groups. In either case, the self-selected nature of the SMPY population makes an evaluation of these findings difficult. Benbow concluded that support for a "primarily environmental explanation" is lacking in this high-ability population, and that a more fruitful search for the causes of "extremely high mathematical reasoning ability" might be conducted in the biological domain (p. 29).

Stanley (1982) offered a hypothesis involving hemispheric dominance. It suggested that girls who score high on the SAT-mathematical sections do so because they are "brilliant verbally," whereas boys rely on "the nonverbal hemisphere of the brain." Dorans and Livingston (1987) submitted this hypothesis to an empirical test, examining the mean SAT-verbal scores and their standard deviations for male and female examinees from the 1982 administration who scored at the level of 600 and above on the SAT-mathematical sections. Dorans and Livingston were searching for higher SAT-verbal scores for females than for males in this group, and less variance in those scores, as support for Stanley's hypothesis. Their results were mixed: females scored higher, but the standard deviations of their scores were lower than those for males. This partial support for the hypothesis might have been a function of the age differences between the Stanley sample and the Dorans and Livingston sample, the former 12- and 13-year-olds, the latter high school juniors and seniors.

Benbow and her colleagues have identified three physiological correlates of high mathematical ability: left-handedness, symptomatic atopic disease (allergies), and myopia (Benbow 1986, p. 29). Because she thinks that left-handedness and allergies may be related to "bihemispheric representation of cognitive functions or the influence of fetal testosterone," the latter two may be additional physiological correlates of mathematical ability. She will investigate these correlates in years to come.

Social and Psychological Explanations

Few would argue that social factors do not play an important role in the cognitive development of individuals. At the same time, many of these factors are ingrained in long-standing patterns of behavior that, despite the efforts of the feminist movement, remain part of the "nonconscious ideology" (Bem and Bem 1976) of sex differentiation in our society. For this reason, it is possible that some of the subtle differences in the life histories of men and women are and will remain unexamined. This section considers some of the vast number of acknowledged influences that differ for males and females, and that have been offered as contributory mechanisms to differences in test performance.

Socialization Processes

Early Sex-Role Development. Sex roles refer to the distinctions based on gender that are made and adhered to by society. Such roles include behaviors that are expected of and rewarded in males and females. There is ample evidence that boys and girls are treated differently from birth (Golden and Birns 1976; Block 1976; and others) and perhaps even before, in an age of increasing knowledge about the gender of the unborn child. Specifically, boy babies are handled more than girl babies, and girl babies are spoken to more often than boy babies (Lewis and Freedle 1973). Parents react more positively toward their toddlers when the children are engaged in gender-appropriate behavior (Block 1976). Moreover, parents' behavior is not always congruent with their stated attitudes, as at least one observational study (Fagot 1978) revealed.

During early childhood, personality differences between boys and girls begin to emerge. Differences have been documented in aggression, activity level, dominance or "toughness," and sociability, traits

that males tend to possess to greater degrees than do females; and empathy and dependency, which boys and girls manifest in distinctly different ways. Differences have also been documented in play behavior, expressed, among other things, in a general preference for outdoor, active play in boys and for indoor play with toys for girls. Whether these are causes or effects of differential treatment by parents, there is evidence that parents react differently to the same trait in boys and girls. For example, parents tend to respond to dependency behavior in girls by encouraging them to stay close and in boys by encouraging them to move away from parents. Boys receive more encouragement for achievement, self-reliance, and competition by both their fathers and mothers (Block 1976). Parents begin training boys for independence earlier than they do girls and emphasize such training more (Hoffman 1977). Boys receive more punishment than girls, and more rewards. Boys' and girls' rooms are furnished differently (Rheingold and Cook 1975), boys' with a greater variety of toys and with more action-oriented equipment. And parents instruct their sons and daughters in the different behaviors expected of them by providing them with different toys: boys' are "moveable and active and complex and social," whereas girls' are "the most simple, passive, and solitary" (Brooks-Gunn and Matthews 1974).

Boys and girls of elementary school age have different leisure-time interests. Boys are more interested than girls in (among other things) guns, team sports, and in making and fixing things. Girls prefer dolls, sewing, cooking, and dancing (Zill 1985). Boys are more likely than girls to be left unsupervised after school, and girls are more likely to be picked up by parents and caretakers (Houston 1983), a circumstance that may curtail the development of risk-taking and exploratory behavior in girls. That from an early age, children understand and act upon the messages and instructions that come from these differences has been demonstrated in studies that employ a variety of measures from self-reported attitude scales through observations of behavior.

Schooling. Boys and girls appear to experience school differently. One manifestation of the difference is the fact that boys initially have more difficulty learning to read. Although by age 10, most of them have caught up, Brooks-Gunn and Matthews (1979) estimate that between three and 10 times as many boys as girls have learning and/or behavioral disorders in school, the most common of which is the failure to read or to read well (p. 174). By way of contrast, math achievement for boys and girls is roughly equivalent throughout the years in elementary school. Toward the end of that time, boys' math achievement exceeds girls' and continues to do so through high school and college. Some early studies (Kagan 1964; Milton 1957) correlated these achievement findings with the sex-typing of the achievement areas: reading is seen by males and females alike as a feminine activity, math as a masculine one. One study of third graders (Schickedanz 1973) revealed that boys who perceived reading as a masculine activity read better than boys who thought it a feminine activity. Houston (1983) demonstrated relationships between children's perceptions of the sex-role appropriateness of different activities (reading, math, and art) and their motivation to achieve in these areas.

In a similar vein, performance in math among elementary school children has been found to be related to their and their parents' ideas about the value of math, which parents, at least, value more for boys than for girls. Other studies of mathematics achievement, beginning with Hilton and Berglund (1974) and including many more recent investigations (e.g., Steinkamp and Maehr 1984; and Chipman et al. 1985; both collections of articles about women and mathematics achievement and participation), demonstrate the reciprocal influence of interest and achievement in math, and of both of these on boys' and girls' different expectancies for success in math.

Boys and girls are treated differently by their teachers. An observational study of second grade teachers (Leinhardt, Seewald, and Engel 1979) revealed that the teachers spent more time teaching reading to individual girls and less time teaching them math. Boys, on the other hand, received less direct instruction in reading relative to girls, and more in math. Despite efforts in recent years to decrease the demonstrated but often unconscious differences in the behavior of elementary school teachers toward boys and girls, recent observational studies by Sadker and Sadker (1985a) have documented differences that remain. Boys receive more attention than girls, both praise and rebuke. Boys are called upon more, given more time to respond ("wait time," or the time a teacher allows before issuing feedback or going on to the next student), and provided with more substantive feedback than girls. Diener and Dweck (1980), building on Weiner's theory of the attributions people make on the basis of feedback about their own achievement, hypothesized and offered evidence that girls use such feedback in a very different manner from boys. Girls tend to internalize feedback about failure and attribute success to external forces (luck, the simplicity of the task). Boys, on the other hand, internalize feedback about both success and failure, attributing the former to their ability or motivation or both. Moreover, based on observations in class-

rooms, Dweck (1978) maintains that teachers treat the successes and failures of boys and girls differently, somehow encouraging boys to try harder and allowing girls to give up. The differences, Dweck claims, lead boys to become more and more self-confident and self-assured about their academic potential, and girls to develop a stance of "learned helplessness," in which they tend to distrust their own efforts as mediators of success and, essentially, don't try as hard.

Educational materials—textbooks, books for "free" reading, and, in recent years, software—often portray males and females in stereotypic ways (*Women in Words and Images,* 1972) and appeal to boys and girls in ways that may enhance rather than eliminate differences in achievement and motivation (Lepper 1985). Although efforts to change these materials have succeeded in improving textbooks and the print materials that support instruction, software has not kept pace with the changes, and many schools, for a variety of reasons, keep old texts and materials even after new ones have been adopted.

Classroom organization, according to some observers (Sadker and Sadker 1986), also favors boys. Teachers tend to encourage individual effort or create instructional groups that compete with one another in the service of learning. Slavin (1978) offers evidence that girls perform better in cooperative learning situations, which are not typically employed in classrooms. Finally, teachers tend to assign chores to boys and girls in stereotypic fashion, tasks requiring strength to boys and housekeeping chores to girls.

Individual Differences

A number of researchers have examined gender differences in test performance as a function of other individual differences that vary by sex. Whether these differences are considered the causes or covariates of differences in test performance, some of the most commonly cited are briefly discussed here.

Cognitive Styles

Cognitive styles refer to individual differences in preferred ways of organizing and thinking about the world (Messick 1984). The best-researched cognitive style is one that describes the degree to which individuals are influenced by objects in their visual field, namely, field dependence or independence (Witkin et al. 1962). Two common methods of assessing field dependence/independence are with the Rod and Frame Test (RFT), the use of which is described elsewhere in this review (see page 26); and with the Embedded Figures Test (EFT), a paper-and-pencil measure in which subjects are asked to remember a simple geometric shape and locate it within a more complex figure. Subjects whose judgments of true vertical in the RFT are influenced by the tilt of the frame that surrounds the rod and who are less able to segregate a figure from its context are classified as "field dependent." Others, who are not influenced by the tilt of the frame and who are adept at separating figure from context, are classified "field independent." In general, females have been found to be more field dependent than males (Witkin et al. 1962). Differences in field dependence/independence have been found to be correlated with differences in problem-solving ability, conformity, and concern with the reactions of others.

Sherman (1967) has argued that sex differences in field independence are an artifact of sex differences in visual-spatial ability. Hyde, Geiringer and Yen (1975) administered the Rod and Frame Test and the Embedded Figures Test to a group of college students, along with tests of spatial ability, arithmetic, vocabulary, and word fluency. Males performed better than females on the RFT, the EFT, and the tests of spatial ability and arithmetic; females performed better on the vocabulary and word fluency tests. Analysis of the data controlling for differences in spatial ability, however, eliminated the sex differences in the RFT, the EFT, and the arithmetic test. Controlling for differences in vocabulary had little effect on the remaining results. Developmental data also show that sex differences in field independence and sex differences in spatial-visual ability tend to co-vary with the age of the subject (Crosson 1984).

Achievement Motivation

Because they did not behave in ways that were consistent with the model he developed to explain individual differences in motivational factors related to achievement, McClelland (1961) dropped females from much of the research he conducted to validate his "need for achievement" construct. Horner (1970), agreeing that achievement motivation differed for males and females, studied the orientation of females to achievement and concluded that young women suffer from what she then labeled "fear of success." Condry and Dyer (1976) reinterpreted the factor previously termed "fear of success" and viewed it as an accurate assessment by achieving women of the difficulties they are likely to encounter. Harter (1983) asserted that males and females have equal motivation to achieve, but that males have greater "mastery motivation." Lenney (1977) concluded that, compared with men, women have lowered expectancies of success in intellectual domains. Compared with males, whose self-confidence is more stable, females' self-confidence tends to vary with social cues and reinforcement. Dweck's work has already been mentioned in the context of differences in the classroom experiences of boys and girls. Her contention

is that girls' "helpless achievement orientation" is responsible for their lower (compared to boys) math achievement, because math is an area in which helplessness is most likely to undermine performance (Licht and Dweck 1983). Consistent with this hypothesis, Wolleat et al. (1980) found significant sex differences in attributions about success in math. Their data led them to conclude that women's lesser confidence and persistence in the area of mathematics may be a function of their attributions of success or failure.

Causal Attributions

Weiner et al. (1971) developed the original theory of attribution related to achievement on which much of Dweck's work is based. In it, he identified four basic causes to which individuals attribute their success or failure in any domain: ability, effort, luck, and the difficulty (or lack thereof) of the task at hand. The general theory, which has been supported empirically (Weiner 1979; Frieze et al. 1982), holds that there are individual differences in the ways in which people make attributions about their successes and failures, and that these are related in systematic ways to expectancies regarding future performance and to achievement. The possibility that males and females may make different attributions for their successes and failures has been promoted as a cause of the differential performance of males and females on tests of achievement and ability. Three additional theories have been proposed to explain sex differences in attributions (Frieze 1980); each of the theories posits a different mechanism for the differences. The first hypothesizes a general externality, in which women tend to attribute both their successes and failures to external causes and consequently to withdraw from achievement situations, at least in comparison with men. A second model hypothesizes a general mode of self-derogation, in which women attribute their successes externally, but their failures to internal causes. In this mode, women are believed to discount positive information about their achievement. The third model claims that women have generally low expectations about achievement and attribute their failures to stable factors and their successes to unstable ones. All three models emphasize the importance of initial expectancies in individuals' reactions to feedback about their performance.

Whitley, McHugh, and Frieze (1986) conducted a meta-analysis of 28 studies that examined sex differences in attributions related to success and failure in an effort to evaluate the support for each of the theories. Their results yielded small effect sizes, only two consistent sex differences, and minimal support for any of the three theories. The consistent differences were that men are more likely than women to attribute their outcomes to their ability, regardless of outcome, and that men are less likely than women to attribute either their successes or failures to luck. The meta-analysis also revealed that the results of any given study are strongly affected by the way in which attributions are measured and by other situational variables like the context of the research and the task domain. These findings raise questions about the generalizability of attributional findings beyond the specific contexts in which they have been measured. In fact, Whitley et al. summarize their findings in this way: "From the research to date, one would be forced to conclude that there is no sex difference in attributional tendencies sufficiently large to explain male and female achievement patterns" (p. 128).

Educational Variables

Differences in Educational Experiences

Several of the differences in the ways in which boys and girls are educated, especially in the early grades, have been noted in an earlier discussion of socialization and the development of sex roles. Evidence from elementary school classrooms suggests that boys and girls receive different treatment and respond differently to such treatment. In a series of case studies of elementary school children, Grieb and Easley (1984, p. 317) identified a double standard in the area of mathematics teaching. This double standard rewards (mainly) white, middle-class boys who are independent and self-confident and, according to the authors, "creative in their study of mathematics." By not confronting their nonconforming behavior, teachers allow them to operate outside the main classroom ethos in mathematics, whereas females and minorities are held to more conventional standards. These standards involve conformity to "the social norms of arithmetic," which conceptualize mathematics as a set of arbitrary procedures to be undertaken in a fixed sequence. In this mode, the teacher typically requires that the student know the algorithm before proceeding with a problem. The model student, under such conditions, follows instructions, memorizes algorithms and number facts, and learns to distrust any understanding beyond that which is presented. The students who are most likely to resist such instruction, and to emerge untouched, are white, middle-class boys, who then develop the independence that the authors claim is required for achievement in higher-level mathematics.

Peterson and Fennema (1985) examined some instructional correlates of high and low achievement in mathematics among students in 36 fourth-grade

classrooms. Students were tested using the NAEP Mathematics Achievement Test in December and again in May, and residualized gain scores were computed separately for boys and girls. Group means for boys and girls were not significantly different at pretest, posttest, or with respect to gains. For purposes of this study, the authors distinguished between high- and low-level test items, compared performance on each for boys and girls, and examined the effects of classroom variables on performance. They found that although student engagement and nonengagement in mathematics activities in the classroom were related to students' mathematics achievement in predictable ways (i.e., engagement was positively correlated with achievement for both boys and girls and nonengagement was negatively correlated with achievement), the global variables of engagement and nonengagement did not adequately explain sex-related differences in achievement. Instead they found that they needed to examine the kinds of activities in which boys and girls were (or were not) engaged and to examine performance on high- and low-level items separately. For example, engagement in competitive mathematics activities was negatively related to achievement on low-level items for females, but positively related to achievement on low-level items for males; engagement in cooperative mathematics activities was positively related to both low- and high-level achievement for girls, but negatively related to high-level achievement for boys. Similarly, engagement in social activities and one-on-one activities with the teacher were negatively associated with achievement on high-level items for girls but had no effect on boys' achievement. These results suggest that classroom dynamics may be related to achievement in complex ways, which could create dilemmas for teachers in their management of instruction.

Patterns of Course-Taking

Considerable attention has been devoted to differences in course-taking behavior as these relate to differences in achievement, particularly in mathematics. One set of explanations of differences in performance on measures of quantitative ability and mathematics achievement is based on the premise that these differences are largely if not totally the result of the fact that females take fewer and fewer higher-level courses in mathematics. Jones (1984), for example, describing gender differences in MCAT scores, concludes that "the historical performance differences between men and women are no doubt related to different interest patterns reflected in course selection during high school and college." Doolittle (1985, p. 1) argues that Differential Item Functioning or Differential Item Performance results can legitimately be regarded as indicators of group differences in preparation or instruction rather than as evidence of test or item "bias."

Using data from HS&B, Ekstrom et al. (1988) chronicled some of the changes in the school experience of high school students during the 10-year period between 1972 and 1982 as part of an effort to explain the decline in test scores found in those data. For both males and females, the mean number of courses taken in each of the "basic" areas of the curriculum decreased over the period, supplanted by vocational education courses, the only curricular area to show gains. Although the large numbers in the sample render even minor changes statistically significant, there are some that stand out. For example, males in 1972 reported taking an average of 4.22 courses in mathematics, females 3.63. The comparable figures for 1982 were 3.88 for males and 3.52 for females, significant reductions in both cases, but significantly larger for males than females. Thus the gap between males and females in mathematics course-taking was reduced. Similarly, average numbers of science courses taken were 3.93 in 1972 and 3.10 in 1982 for males, and 3.48 in 1972 and 2.86 in 1982 for females, a larger reduction for males than for females.

Attempting to account for data from the Women in Mathematics Project and 1977–78 NAEP data that showed gender differences in mathematics achievement among twelfth graders that had not been apparent among ninth graders, Armstrong (1981) examined patterns of course-taking. The Women in Mathematics Project was an investigation designed specifically to address the question of sex differences in the development of mathematical skills. The main study included over 375,000 students in 987 schools chosen to be representative of American public and private secondary schools. Armstrong noted, at that time, that the large sex differences in participation found in earlier studies no longer existed; that both the NAEP data and the Women in Mathematics survey data showed few differences in participation for general math courses, algebra 1, and geometry; and that both surveys found statistically significant differences favoring males for different advanced mathematics courses (algebra 2 and probability and statistics in the Women in Mathematics Project and trigonometry, calculus, and precalculus in NAEP). However, at all levels, even the differences that were not statistically significant favored males.

Armstrong examined achievement within different levels of participation and found that men at nearly every level had an advantage in solving word problems. She concluded that achievement differences were not solely a function of differences in course-taking. Nor did sex differences in achievement appear from these data to be related to differences in spatial visualization. Armstrong (1981, p.

369) concluded that perhaps the sex differences in achievement are the result of "differential learning and practice of mathematics outside of school," to the choice of different problem-solving strategies by men and women, or to personality variables like motivation, perseverance in solving problems on tests, and self-confidence in mathematics.

Wise (1985) examined a subset of the data from the same Women in Mathematics Study and derived somewhat different conclusions. Wise's special subsample included 7,500 of the total group who were tested as ninth graders in 1960 and again as twelfth graders in 1963 to examine factors influencing math gains during high school. In the ninth grade, there was a small (.07 standard deviation) difference favoring males in mean mathematics achievement; however, male gains in math achievement during high school were more than twice the size of female gains, increasing most sharply after the tenth grade. The strongest predictors of twelfth-grade math achievement were ninth-grade math achievement (r=.78) and the amount of math taken in high school (r=.73). In fact, higher ninth-grade scores were associated with higher raw-gain scores, demonstrating that individual differences did not remain constant but increased during that period. After controlling for amount of math taken, Wise found that sex differences in achievement were virtually nonexistent. Acknowledging that females who took advanced math courses in high school were a more select group than males with the same level of participation, Wise controlled for ninth-grade achievement and found that females scored .1 standard deviation lower than males. Based on these data, Wise concluded that roughly seven-eighths of the relationship between sex and twelfth-grade math achievement could be attributed to math courses taken and achievement differences in the ninth grade. Wise identified three additional factors that predicted gains in math achievement: general academic aptitude, interest in math and math-related occupations, and low levels of participation in extracurricular activities. Wise also noted that, in this sample, sex differences in career interests and interest in math itself were already evident by the ninth grade. These differences predicted sex differences in the number of courses taken and in math achievement during the high school years.

Armstrong (1985) used data from her 1978 survey of samples of 13-year-olds (n=1,452) and twelfth-grade students (n=1,788) to examine a number of factors hypothesized to be related to achievement and participation in mathematics. Her 90-minute survey included measures of mathematics achievement and participation; sex-role stereotyping; career and academic plans; attitudes toward mathematics; parental influence; influence of others; and several background variables. Armstrong then compared her results with results obtained in the 1977–78 National Assessment of Educational Progress (NAEP) in mathematics. (These findings were reported earlier in this section). Both of the studies showed patterns of performance that were similar to those reported by Maccoby and Jacklin, in which 13-year-old males and females demonstrated approximately the same level of mathematical understanding and skills (in fact, 13-year-old females were better at computation and spatial visualization than their male counterparts) but males caught up with females and even surpassed them in certain areas of mathematics as high school seniors. Among seniors, there were no sex-related differences for computation or algebra but large differences favoring males in problem-solving. There were few sex differences in participation in lower-level high school mathematics (more females took business or accounting mathematics and more males took algebra II and probability and statistics), but significant differences reported by 17-year-olds and high school seniors in enrollment in trigonometry, precalculus, and calculus. At the same time, these differences were smaller than the differences reported in earlier studies, suggesting that the disparity in course-taking behavior by males and females might be diminishing. Armstrong identified three groups of variables with the greatest effect on participation in higher-level mathematics courses: positive attitudes toward math; perceived need for and usefulness of math; and positive influences of parents, teachers, and counselors.

Benbow and Stanley (1980, 1983) concluded that differential course-taking does *not* account for sex differences in mathematical ability, based on data from the Study of Mathematically Precocious Youth (SMPY) collected over an eight-year period. The SMPY population included about 10,000 students in grades 7 through 10 who, it will be recalled, had qualified for inclusion in the study by scoring among the upper 2, 3, or 5 percent "in mathematical ability as judged by a standardized achievement test" (Benbow and Stanley 1980, p. 1262), and who then took the SAT. Their SAT-mathematical results showed a mean difference of about .5 standard deviation favoring males, with greater disparities at the upper-score levels. Because the SAT-mathematical section was administered to these students before they started to diverge in terms of number and level of mathematics courses taken, Benbow and Stanley concluded that course-taking in mathematics could not alone explain the difference in test scores.

Thus, although the discrepancy between males and females in participation in mathematics in high school has diminished over recent years, males still

appear to take more math courses than females, particularly at the higher levels. Moreover, within those courses, males seem to outperform females on measures of mathematics achievement. These data continue to concern researchers and policymakers alike.

Eccles (Parsons) et al. (1983) combined both cross-sectional and longitudinal data in an effort to model the factors that affect differential participation in mathematics. Her study of 339 students in grades 5 through 11 included parents and math teachers as well as data from multiple sources: student records; questionnaires to students, teachers, and parents; and classroom observations. The questionnaires for students included a range of attitudinal and self-report measures related to aspirations, sex-role identity and perceptions, patterns of causal attributions, and perceptions of parents' and teachers' beliefs about them, the students. Parents were asked about their own attitudes and those of their children. And teachers were asked for their beliefs about the causes of sex differences in participation in mathematics and for judgments of each child's math ability and performance. Teacher–student interactions were observed for 10 sessions in each of 18 mathematics classes. A control group of 329 students was added during the second year of the study. A variety of analyses, both descriptive and relational, were performed on the data, culminating in a series of cross-legged panel analyses to test causal inferences.

Eccles's results were summarized in a path-analytic model (see Appendix B for the model) that implicates parents and teachers in the attitudes that students have toward mathematics and, therefore, in their patterns of course-taking in high school

Instruction in Specific Skills

There is some evidence that instruction in areas in which females have traditionally been regarded as inferior can reduce or eliminate gender differences. In a study of 1,364 students in 74 high school classes, Senk and Usiskin (1983) were able to develop equal facility among males and females at writing geometry proofs. The subjects in Senk and Usiskin's demonstration ranged in age from 14 to 17 and attended schools that were chosen to represent a national cross-section of educational and socioeconomic conditions. The authors characterize geometry proof-writing as "a high level cognitive task," asserting that it is "considered among the most difficult processes to learn in the secondary school mathematics curriculum" (p. 188). Subjects were given a test for entering knowledge of geometry terminology and facts and, at the end of the school year, a standardized geometry achievement test and one of three forms of a proof test devised for the project. Females scored significantly lower than males on the pretest, the scores of which were used to adjust the proof posttest scores. With these adjustments, total scores for females were higher than total scores for males (significantly so for one of the forms), and mean number of proofs correct was similarly higher for females than males. The authors examined these results for three selected subsets of their population: the top-scoring students on each form of the test; a set of seventh and eighth graders who were accelerated at least two years in mathematics; and a group of those in the sample who scored in the top 3 percent according to national norms, a group considered comparable to the Benbow and Stanley's SMPY group. In all three groups, Senk and Usiskin found equivalent performance in proof-writing by the identified high-achieving boys and girls.

An effort to improve the visual-spatial skills of junior high school students (Connor and Serbin 1985) showed that at least two such skills—spatial orientation and visualization—could be enhanced by brief training sessions. No sex differences were found in trainability, and there was some suggestion that students who performed relatively poorly on visual-spatial tasks improved more as a result of training than students who performed well.

Integrative Models

Recognizing that sex differences in cognitive abilities are most likely to reflect a complex pattern of influences that operate throughout the lives of individuals, a number of students of the topic have attempted to describe the pattern in a way that respects its complexity. Within the past decade, several models of the development of sex differences in cognitive abilities have been proposed. These models attempt to integrate the findings from the biological, psychological (individual difference), and social domains that have tended to exist in isolation from each other, and to take account of both cross-sectional and longitudinal data. The models tend to share an underlying set of assumptions about 1. the complexity of the process, 2. the likely mutuality of influences, and 3. the simultaneous or sequential contributions of biological factors, individual differences, and socialization processes to whatever is considered the outcome (test scores, mathematics course-taking, career choices). The models also tend to acknowledge the possibility that any given outcome (test scores, for example) might itself contribute to another outcome (like career choice).

Such models have been developed by Ethington and Wolfle (1986), Farmer (1987), Boswell (1985), Kavrell and Peterson (1984), Lockheed et al. (1985), Eccles (Parsons) et al. (1983), Stallings (1985), and

Wise (1985). The models are typically based on different data sets and therefore vary with the data they are attempting to explain. They vary with respect to predicted outcome (for example, for Farmer, predicted outcome is career and achievement motivation; for Kavrell and Peterson, it is "cognitive performance"; for Wise, mathematical performance; for Eccles, academic performance; and for Lockheed et al., it is academic performance in mathematics, science, or computers); in the ages on which they focus (Wise, for instance, concentrates on twelfth graders, Ethington and Wolfle examine data for tenth and twelfth graders both cross-sectionally and longitudinally, and Lockheed et al. look at middle school students); and in the explanatory variables they include (Kavrell and Peterson, for example, include biological factors whereas the others do not; Stallings and Eccles include classroom interaction data). Nonetheless, the models represent attempts to deal with the phenomenon of gender differences in a multivariate fashion. (See Figures 1-6 in Appendix B.)

Ethington and Wolfle, for example, used data from the first follow-up of the 1980 sophomore cohort of the HS&B study and created a latent-construct model of the process of mathematics achievement. The data included measures of mathematics and verbal ability, mathematics achievement, and exposure to and attitudes toward mathematics. Having constructed the model, Ethington and Wolfle compared the process for males and females and found that it differs for the sexes, and that it is probably more complex than prior research suggests. The factors in the model with positive effects on mathematics achievement—higher math ability and more positive attitudes toward math—led to greater increases in math achievement for men than for women. And high verbal ability led to greater exposure to mathematics for men than for women. The factor with the highest *negative* effect in the model—verbal ability on attitudes toward mathematics—had a stronger negative influence for women than for men. Ethington and Wolfle concluded that "questions about average male–female differences in mathematics achievement have little meaning unless the question is asked in relation to specific values of prior ability and educational experiential variables" (p. 73).

It can be seen from the single example given that the models do not lend themselves to easy summarization, but schematic representations of several are included in Appendix B. Their usefulness lies in part with their acknowledgement of the interactivity of the variables included and their attempts to apply appropriately complex approaches to what is clearly a complex issue. From the perspective of developing strategies for intervention to eliminate or minimize sex differences, the models are helpful in that they identify appropriate targets for intervention (e.g., training in spatial skills, the classroom behaviors of teachers, and the attitudes of parents). In fact the authors of several of the models (e.g., Eccles 1983; Stallings 1985) offer a range of suggestions for intervention based on their findings. The models are also useful in the generation of new hypotheses to be tested in controlled laboratory studies. In short they offer approaches to the integration of existing data and collection of additional data.

Demographic Explanations of Trends

Because of their potential impact on decisions about admission, placement, and scholarship awards, sex differences in admission tests are particularly vexing to those concerned with educational equity. And because of its prominence and visibility as an admission criterion, the SAT has received the lion's share of critical attention where sex differences and trends in these have been concerned. So seriously was the recent decline in SAT scores regarded that a national commission was established to investigate its causes. The commission collected 79 different hypotheses about the decline, among them television, poor training of teachers, watered-down textbooks, drugs, parental neglect, nuclear testing, and food additives (Wharton 1977).

One explanation of the score vicissitudes (and prediction of future trends) comes from the confluence model, a theory that explains score trends by relating them to changes in family patterns (Zajonc and Bargh 1980 a and b; Zajonc 1986). According to the confluence model, the intellectual environment of the family has a significant effect on the mental growth of its children, an influence that changes with the size of the family, the spacing of children, and their relative position within the family. The model is written in terms of individual intellectual growth curves but can be extrapolated to aggregate data. In fact Zajonc explained the SAT decline in 1976 in terms of the fact that cohorts taking the SAT between 1963 and 1980 came from families whose average size increased steadily over the years. In a more recent analysis (Zajonc 1986), he increased the predictive power of the model by incorporating the proportion of seniors who take the test. Using two factors—birth order and the proportion of those born who take the SAT—Zajonc claims to have accounted for 67 percent of the total variance in SAT trends. The relevance of Zajonc's analysis to the issue of sex differences is his observation that the trends in proportions of seniors taking the SAT are different for men and women. These, in turn, result in different standardized coef-

ficients for the effects of the proportion of seniors taking the SAT and for birth order when multiple regression analyses are performed on the data. According to this model, the proportion of men who took the SAT between 1973 and 1985 was determined to a large extent by their order of birth. However, birth order was less influential a factor in the likelihood of women taking the test. Inexplicably, the two factors accounted for only 44 percent of the variance in men's average SAT scores, but for 78 percent of the variance in women's, differences that Zajonc claimed his data are insufficient to explain.

Paulhus and Schaeffer (1981) found support for the confluence model for males but not for females. In their study the number of older siblings was negatively associated with SAT scores of both male and female college students, but number of younger siblings appeared to be negatively associated with SAT scores for males and positively associated with SAT scores for females. This finding was not supported in later research by Steelman and Marcy (1983) with a larger, more nationally representative group of students, using an IQ measure rather than SAT scores. Instead their results showed a difference by domain: females' verbal IQ scores were less likely than males' to be negatively associated with number of siblings, but their nonverbal IQ performance was more likely to be impaired by larger numbers of siblings. It is difficult to interpret these findings, much less make sense of the contradictions among studies. They do, however, demonstrate that the factors associated with test performance may work differently for males and females.

Burton (1987) examined trends in several of the voluntary testing programs, including the SAT, mainly as part of an effort to explain the decline of women's SAT-verbal scores. (This difference, although of relatively "small" magnitude—about .12 standard deviation—and very slight practical significance, evoked considerable concern on the part of both the test sponsors and the general public.) Burton observed that for the SAT as well as several other college and graduate school admission tests, the relative proportion of women taking the test increased as the relative performance of women compared to men on verbal tests declined.

Burton, Lewis, and Robertson (1988) explored the role of demographic changes in the decline of the SAT-verbal scores using samples of test-takers from 1975, 1980, and 1985. Using multiple linear regression techniques, Burton et al. examined the contributions of gender, ethnic group membership, socioeconomic status, high school course-taking, and proposed college major. The analyses established that women who take the SAT are, on average, different from men and that the background differences between men and women are significantly related to score differences. Burton et al. interpreted their results as meaning that, were the men and women who take the SAT more alike in background, women's SAT-verbal scores would be at least as high as and perhaps higher than men's. Although the mathematics difference would not be totally eradicated, were women SAT-takers more like men with respect to the background characteristics studied, the score differential would be reduced by about half. At the same time, the analysis did not account for the declining trend in women's scores compared with those of men.

Citing earlier analyses that had demonstrated that the downward trend was not attributable in large measure to changes in the test (Burton 1987) or to individual items (Wendler and Carlton 1987), and the fact that the downward trend is also reflected in a range of different verbal measures, Burton et al. concluded that the SAT trend is most likely due to changes in the education of women.

Ekstrom, Goertz, and Rock (1988) used two different analytic approaches in their examination of trends in test scores of the national samples of students who took part in the National Longitudinal Study (NLS) of 1972 and High School and Beyond (HS&B) in 1980. The first analysis partitioned the mean test score changes by population changes; the second employed an analysis of covariance. The first analysis considered the amount of change attributable to the numerous demographic changes that occurred between 1972 and 1982 in the makeup of the population of high school seniors represented by the data. The 1982 group included greater representation of minority groups, Southerners, students in non-Catholic private schools, and nonacademic curriculums. Except for the last difference, Ekstrom et al. showed that declines in the reading, vocabulary, and mathematics test scores were more likely to be due to changes within the groups in question than to their relative representation in the test population (p. 79). In this analysis, girls showed larger declines than boys on the verbal tests—reading and vocabulary—but less decline compared to boys in mathematics. During this period, it will be recalled, the differential in favor of boys with respect to mathematics course-taking was significantly reduced.

A second analysis examined the impact of selected blocks of variables controlling for other, confounding variables. The results of this analysis showed that the primary contributor to the score declines in all the tested areas was students' school experiences. Demographic changes and school characteristics contributed relatively little to the score declines when contrasted with the impact of changes in school experiences during the ten years in question. Of the four school experience variables that

appeared to contribute most to the declines, taking fewer semesters of foreign language courses, spending less time on homework, taking fewer semesters of science, and not being in the academic curriculum appeared to do the most damage. With respect to the reductions in both foreign language course-taking and time spent on homework, females fared worse than males. That is, the decline in foreign language course-taking and in time spent on homework was greater for girls than for boys.

Ekstrom et al. also conducted analyses of achievement gains in high school by examining the sophomore and senior results in the context of other variables. Number of language courses taken was one of the two largest predictors of gains in both vocabulary and reading. Amount of homework done was also an important predictor of reading gain. The major determinant of gains in mathematics achievement was the number of mathematics courses taken; sex is the next largest determinant, with males gaining approximately 1.0 score points more than females (p. 101). In science achievement, similarly, males gained more than females, and the number of science courses taken was the second largest determinant of achievement gains in the area. Gains in writing were associated with being female (females gained about 1.5 score points more than males) and with absence of discipline problems, taking language courses, and doing more homework.

Characteristics of the Tests Themselves

Critics of the SAT have asserted that the tests themselves contribute to gender differences in performance. At least one study prior to the current flurry of activity in the area of test and item bias found that females performed less well on items with "male" content and better on items with "female" content (Donlon, Ekstrom, and Lockheed 1979). A recent study by a group concerned with fairness in testing (Loewen, Rosser, and Katzman 1988) examined the performance of 1,112 students in a coaching class on one mock form of the SAT. They identified 17 items—7 verbal, 10 math—that favored one sex or the other and concluded from simply examining the items that male-oriented vocabulary in both the verbal and math items may have adversely affected females' performance.

A more precise technique for examining the relative performances of males and females while minimizing its confound with differences in other, related factors (test score or patterns of course-taking, for example) is the analysis of differential item performance (DIP) or differential item functioning (DIF). This technique focuses on differences in item-level performance for groups that are comparable on some dimension (e.g., ability or course-taking). The technique identifies items that function differently for members of different groups (men and women, or black and white examinees). The items so identified can then be examined for content or form that may favor one group or another, or compared with other items in a particular test. The analyses are particularly useful as sources of hypotheses regarding the differences observed, hypotheses related to the form or content of the items or to the cognitive process required to respond to them. DIF or DIP analyses have been conducted for various ACT examinations (Doolittle 1985 and 1987; Doolittle and Cleary 1987; Welch and Doolittle 1988); for the SAT-V (Lawrence, Curley, and McHale 1987; Wendler and Carlton 1987; Carlton 1987) and the SAT-M (Dorans 1982); the National Teacher Examination (NTE) (McPeek and Wild 1987); the GRE and GMAT (Wild and McPeek 1986; Pearlman 1987); and for NAEP (Hudson 1986). Several of these studies are summarized below.

Doolittle and Cleary (1987) examined differences in performance on a special form of the ACT-mathematics test among high school students with similar records of mathematics course-taking. (Interestingly, even their careful matching of test-takers on the basis of course-taking patterns did not totally eliminate the male–female differences. For example, 18 percent of the females selected for participation in the study but 23 percent of the males reported having taken introductory calculus.) Using six types of items—arithmetic and algebraic operations, arithmetic and alegebraic reasoning, geometry, intermediate algebra, number and numeration concepts, and advanced topics (trigonometric functions, combinations and permutations, probability and statistics, and logic)—the authors found that geometry and arithmetic and algebraic reasoning problems tended to be relatively more difficult for female test-takers and that intermediate algebra and arithmetic and algebraic (algorithmic) operations problems tended to be relatively less difficult for them. In their discussion of these findings, the authors suggest that the primary feature distinguishing these groups of items, particularly the operations and reasoning problems, is context. The computation items involve explicitly described operations, whereas the latter are mainly word problems that demand that the test-taker develop an appropriate strategy for solving the problem before carrying out the required operations.

DIF studies based on the SAT match students on the basis of their SAT scores rather than on the basis of patterns of course-taking. In a review of such studies, Dorans (1982) found that items classified as "regular math," which, incidentally, is a format used

for the full range of mathematics content, seemed easier for women than other item types. However, Dorans found few extreme differences in the performance of even these items.

In a *post hoc* analysis of quantitative items from the Graduate Record Examination (GRE) General Test and the Graduate Management Admissions Test (GMAT), McPeek and Wild (1987) studied the relationship between differential item functioning (DIF) and a number of different variables, matching subjects on the basis of total quantitative score. (On both tests, men scored approximately one-half of a standard deviation higher than women.) On the whole, few of the variables studied showed significant differences between males and females; however, some patterns did emerge.

In the GMAT data set, item type, context, presence or absence of variables, and the quantitative content of the item appeared to be related to male-female differences. Women performed better than matched males on data-sufficiency types of items and less well on standard five-choice questions; consistent with the Doolittle and Cleary (1987) findings, better on items presented in pure mathematical contexts and less well on word problems; better on algebra and less well on geometry items; better on items that contain variables; and better on items that require algebraic manipulation and the calculation of factors and multiples, and less well on items involving ratios and proportions. In the GRE data set, consistent with the GMAT results, women performed better than matched men on algebra items and less well on geometry items; and better on questions that contained variables. Unlike the patterns observed for the GRE, women represented in the GMAT data set performed less well than matched males on questions that could be solved more easily by estimation than by computation.

In a first-pass analysis of differential item functioning of items from three forms of the GRE and GMAT examinations administered in October 1984 and April 1985, Wild and McPeek (1986) identified small numbers of items that appeared to operate differently for males and females. (Overall, males' scores were about one-half of a standard deviation higher than females' on the quantitative sections of both tests; females' scores were insignificantly lower on the verbal section of the GMAT and insignificantly higher on the GRE General Test.) Interestingly, among the very small numbers of verbal items so identified, there were items that favored males and items that favored females in about equal numbers. In particular, reading comprehension questions based on passages with science content appeared to favor men, whereas questions based on passages with humanities content favored women.

More fine-grained analyses of the differential functioning of verbal items from the Scholastic Aptitude Test (SAT) were performed by Lawrence, Curley, and McHale (1987) and Wendler and Carlton (1987). In both cases, subjects were matched by verbal score on the test. Once again, the major finding was of the limited occurrence of items with extreme DIF values. Across the four forms examined in the Lawrence et al. study and the three forms in the Wendler and Carlton study, few (74 of 255, or 29 percent in the Wendler and Carlton study) exceeded the range between −.05 and +.05, and fewer still (9 or 4 percent) exceeded the range—between +.10 and −.10—considered problematic. Those items that *did* exceed this range tended to be discrete items, to appear in longer (45-item as opposed to 40-item) sections, and to be based on science content. Among items measuring reading comprehension, those with science content were more difficult for women than those with humanities content, and science passages based on content considered technical (as opposed to content reflecting the history or philosophy of science) were generally more difficult for women. Among discrete items, analogies seemed the most problematic for women, who again performed better on items in the realm of human relationships whereas men performed better on items in the realm of practical affairs and science. Among sentence completion items, those with "true science" references tended to be more difficult for females than those with "surface" science or no science references in all four of the forms studied by Lawrence et al.

A similar analysis of five forms of the GRE (Pearlman 1987) identified 40 items (of a total of 380, about 10 percent) with extreme DIF values, 15 favoring women and 25 favoring men. Within this group of items, three categories of "discrete" verbal items, and within these, verbal analogies, were disproportionately represented; among the four content categories to which each item is assigned, the preponderance of those favoring men were classified as science or the world of practical affairs. By way of contrast, more items classified as aesthetic/philosophical or as dealing with human relations tended to favor women. Pearlman concluded that item content is one source of differential performance (reflecting, possibly, differential course-taking and different experiences in the physical world), and that the nature of the verbal analogy may operate in yet-to-be-determined ways that systematically favor or disadvantage groups of test-takers.

Welch and Doolittle (1988) examined the relationship between characteristics of items from the ACT English Usage Test, one of its four tests of educational achievement, and gender differences in performance. They found an overall tendency for

female examinees to outperform male examinees on this test, given comparable coursework; however, they found no evidence of differential item performance in the items based on the ACT five-way classification system that distinguishes among punctuation, grammar, sentence structure, diction and style, and logic and organization. Nor did they find support for their hypothesis of a possible advantage for females in algorithmic English usage items and one favoring males in reasoning-oriented items.

Two forms of each of three tests (Communications Skills, General Knowledge, and Professional Knowledge) of the National Teacher Examination (NTE) battery administered between 1983 and 1985 were analyzed for differential item performance (McPeek and Wild 1987). Across the three tests, a total of 61 items, about 9 percent, showed differential difficulty by gender, again about equally divided between those on which women performed better and those on which men performed better. In the Communications Skills Test, men performed better than women on questions based on reading passages about science, an interesting finding in that the questions in this test are not supposed to require any outside knowledge of subject matter. The authors interpreted this finding as a context effect. In the General Knowledge Tests, consistent with many of the studies of male–female differences, males performed better than females in most areas of science. However, females performed better than males on biology questions, and males performed better than females on chemistry and physics questions. In the literature and fine arts sections of these tests, women performed better than men on questions about the performing arts, and men performed better on questions about architecture. In the Professional Knowledge Tests, women performed better than men on questions about interactions among teachers, parents, and students, and less well than men on questions about the legal and organizational aspects of education and about controversial topics in education.

Finally, in a series of analyses, Hudson (1986) examined each of the 57 items from the National Assessment of Educational Progress (NAEP) mathematics assessments of 17-year-olds in 1977–78 and 1981–82. Her data are based on nationally representative samples of more than 2,000 students in each of the years represented. Her results showed that in both years the test was more difficult for girls than for boys with equivalent mathematics backgrounds. In 1977–78, she found 12 items that were significantly biased against females and two that were biased against males; in 1981–82, eight items that favored males and one that favored females. There was no discernible pattern in the content of these items. Hudson then examined a number of variables that she hypothesized might help to explain the sex differences. She found no influence at all for difficulty of the previous item or for item format or for cognitive process involved. She did find significant relationships to gender differences for item difficulty, item discrimination, familiarity (of the problem type), and item content. With respect to the last named, females performed less well than males on items dealing with numbers and numeration, measurement, geometry, and graphs and tables. Hudson also performed a distractor analysis of the items in question and found large sex differences in choice of distractors. Females were more likely to use the "I don't know" option than males who failed to answer the questions correctly.

Using an additional sample of high school students, Hudson performed a protocol analysis of the problems for which she had found sex differences, asking students to articulate their approaches to the items. From this she concluded that males and females thought about the problems quite differently, and that females responded in ways that are consistent with learned helplessness to items that they found difficult. They were less willing to guess, used the "I don't know" option when it was available, and gave up more easily than males.

Of the differential item performance studies cited, only the ACT investigation and Hudson's follow-up study using the NAEP items involved an experimental form or administration of the test. The SAT, GRE, and GMAT analyses were all performed on data from actual administrations of the measures. The authors of the latter studies qualify their conclusions by noting that there are potential confounding factors (like speededness of the tests and location of the items or passages) that may affect differential performance. These factors need to be examined in the context of experimental studies that control for possible confounding variables. Such studies might examine the interactions of various factors and also address some of the hypotheses suggested by examination of the patterns of less-than-significant differences in DIF values (Carlton 1987): that women perform better on questions that test abstract (as opposed to concrete) ideas, and on items in which there is more context (sets of items based on reading passages as opposed to discrete items and on longer as opposed to shorter passages), and worse on items with a strongly negative tone. With respect to mathematics items, NAEP results show superior male performance on those that reflect "higher-order processes."

Doolittle (1984, 1985, 1987) has been engaged in DIP analysis of the ACT Mathematics Usage Test. He has consistently found systematic differences

between males and females in their performance on the items in this test. Across all forms examined, matched by high school course background, females performed less well relative to males on "strategic" items than on "algorithmic" items. In an effort to learn more about the influence of course-taking on test performance, Doolittle (1984) examined DIP among a random sample of 2,669 college-bound high school seniors from a 1983 administration of the ACTM. The mean scaled score in this group was about .5 standard deviation higher for males than females. Males in the group averaged more semesters of math coursework (7.2) than females (6.6) during their four years of high school, and a higher proportion of males reported taking advanced or accelerated math courses (37.1 percent compared with 28.7 percent of the females). Doolittle conducted DIP analyses based on level of course background in mathematics, on gender, and on the two combined (high background and gender). These analyses identified 16 (40 percent of the total of 40) items that were identified with DIP in the course background analysis and 12 (30 percent) in the gender analysis. Strikingly, all the items that were identified for both analyses differed with respect to the direction of the DIP. The gender analysis was repeated, controlling for background at the high level, and the results approximated those obtained in the overall, uncontrolled gender analysis. Among the items with significant DIP, word problems tended to favor the group with fewer math courses and more abstract, intermediate algebra items tended to favor the group with more math courses. Consistent with almost every other study of mathematics performance cited here, geometry items and word problems favored males. Doolittle concluded that his study provided additional support for the idea that gender-based differential item performance in mathematics is not a simple consequence of group differences in mathematical background, even though gender and background interact to influence test results. The results may reflect differences in instruction that are established before high school, or they may reflect specific differences in quantitative skills.

Chipman (1988) examined word problems, an item type that has been demonstrated in many of the studies reviewed here to favor males consistently, from a cognitive standpoint; that is, with respect to the processes thought to be involved in solving such problems. Reviewing earlier attempts to analyze arithmetic and algebraic word problems from a cognitive-processing point of view, Chipman summarized what is known about how such items are solved but concluded that little or no research exists on the influence of problem content on the solving of mathematics word problems. Research on the influence of problem content on performance in logical-reasoning tasks, however, seems to suggest that content may make a big difference.

Specifically, familiarity with the content may make it easier to build a mental model of the situation described in the problem, one of the processes thought to characterize success in solving such problems. Chipman hypothesized that "because the situations that may appear in word problems are life situations which may differ in familiarity for males and females, this is one possible source of sex differences in performance" (p. 16). To test this hypothesis, Chipman reanalyzed the results of an earlier study in which 185 male and 148 female high school students had been given a test consisting of 78 word problems that had previously been rated for relative familiarity of content to males and females. Chipman examined the difficulty of the items separately for males and females in relation to the item characteristics and found that "the sex-typing of item content made a whopping difference to student performance, for both females and males" (p. 24). Females performed much better on the "female" items and much worse on the "male" items. Males performed slightly better on the neutral items. All the items rated masculine were more difficult than average for both sexes, and most of the items rated feminine were less difficult than average for both sexes. However, when she tried to create items experimentally that worked as the existing test items had, Chipman was unable to do so.

All the item-level studies summarized here offer fruitful ideas for continued research, particularly for studies that control some of the confounding factors, like speededness, item content, and degree of context provided. A related approach suggested by the Hudson (1986) study is protocol analysis, in which test-takers are asked to "think aloud" about how they solve problems in an effort to make their strategies apparent. Since several studies (e.g., Hudson 1986; Chipman 1988) have suggested that men and women may have different strategies for solving problems, protocol analysis may provide a useful approach to the differences. One limitation of the usefulness of protocol analysis stems from Nisbett and Wilson's (1977) data that demonstrate individuals' lack of complete access to their own cognitive processes. Nonetheless, protocol analysis could prove helpful in conjunction with other measures.

SUMMARY

Since the publication in 1974 of Maccoby and Jacklin's volume on gender differences, considerable attention has been accorded gender differences in performance on measures of verbal and quantitative abilities. Maccoby and Jacklin claimed, in 1974, to have documented differences in verbal perform-

ance favoring women, and differences in quantitative performance favoring men. Although many of the studies they reviewed, using test results from the 1960s and 1970s, supported these conclusions, more recent studies, particularly meta-analyses, have challenged the conclusions that women sometimes outperform men verbally and that men consistently outperform women in the quantitative domain. Recent studies using test data from the 1980s and earlier have added to the growing body of evidence concerning the relative performance of males and females on a variety of tests.

This review has considered data from a wide variety of sources and testing programs in an effort to describe differences in test performance between males and females and to assess possible causes of such differences. The data came from four major sources: undergraduate, graduate, and professional school admission tests; validity studies; tests using nationally representative samples; and studies of performance at the item level. The four sources provided different kinds of information, each with its own advantages and limitations. Results from admission tests are a major source of data, but they are not representative of the general population, since the students who take such tests are a self-selected group. Testing programs based on national samples *are* representative of the larger population, but because the sample size is often quite large it is difficult to analyze the results in any meaningful detail. Validity studies use test scores to predict the performance of men and women in school or some other setting. The tests used to measure performance also vary. Some, like the College Board Achievement Tests, subtests of the American College Testing Program Examination, and the National Assessment measures, assess achievement in particular domains, like reading, writing, and mathematics. Others, like the SAT-verbal sections and SAT-mathematical sections, and such measures of specific skills as the Shephard mental rotation task, purport to measure abilities or at least long-range achievement that transcends particular curricular emphases. Together, all these sources provide a broad base of information.

Against a backdrop of generally declining performance, the data examined suggest that the differences reported by Maccoby and Jacklin (1974) have become smaller over the past two decades. With respect to performance on verbal measures, the slight historical advantage enjoyed by females appears to have disappeared; with few exceptions, the SAT being the most prominent, male and female performance on verbal measures is virtually identical.

In the quantitative domain, there are still differences favoring males for some measures and some tasks. Males typically score higher on the math sections of many admission tests, on the math tests administered as part of High School and Beyond, on the mechanical reasoning section of the Differential Aptitude Test, and in the math assessment done by the National Assessment of Educational Progress, at least among 17-year-olds. Men also appear to do better on tasks that measure some spatial abilities and on tests of mental rotation. These differences first appear during the junior high school years (although not in the National Assessment of Educational Progress sample) and increase during high school.

The undergraduate admission tests examined in the course of this review included the Scholastic Aptitude Test (SAT) and the Test of Standard Written English (TSWE); the American College Testing Program Examination (ACT); the various Achievement Tests of the College Board; and the Advanced Placement (AP) examinations. On all the tests of quantitative and math ability and achievement, men outperformed women, often significantly. The differences, on average, appear to be larger for tests that claim to measure ability than for achievement tests. There is also a disproportion in the number of males at the upper score levels of the SAT-mathematical sections. However, in recent years the gap has been closing as women's scores have improved. On the SAT-verbal sections, women once had a (relatively small) advantage over men. In recent years men have slightly surpassed women as men's scores have increased while women's have decreased. On other tests of verbal abilities and on achievement tests thought to involve verbal skills, women either show a small advantage over men or score virtually the same as men. Data from graduate and professional school admissions tests show similar patterns. Hence these data support the hypothesis that men are better at quantitative tasks but do not support the hypothesis that women enjoy a consistent advantage over men at verbal tasks. It is interesting to note that tests in specific fields, like the Advanced Placement subject tests or the Graduate Record Examination subject tests, show men doing better in traditional "male" areas like science. It is also interesting to note that, with respect to the subject tests, those taken by more men than women showed men outperforming women. Even the small differences reflected in these data can affect the educational opportunities offered men and women since the tests are used by colleges and universities and other institutions as important bases for decisions about admission, scholarships, and awards. This problem is particularly vexing in the case of the SAT-mathematical sections, since more males than females are represented at the higher scores in the distribution.

Validity studies generally compare the admission test scores (SAT, GRE) of various groups with their first-year grade-point average. Such studies generally find women's test scores to be underpredictive

of their performance and mens' overpredictive. These studies also show women's test scores to be more strongly correlated with and more predictive of performance measures than men's. Validity studies need further review, review that takes into account such potentially confounding variables as differences in the way men and women are graded in their courses and differences in their selection of courses.

In general, researchers disagree about whether there are important differences in males and females in verbal ability. The technique of meta-analysis, which examines effect sizes across many studies, has found small differences in some areas of verbal performance favoring women.

The differences in quantitative ability are larger in magnitude and better supported by the literature. In summary these differences favor males, beginning in junior high school. Boys appear to be better at tasks involving reasoning, and therefore at higher math; girls appear to be better at computational tasks. Some have related this difference in quantitative ability to differences in spatial ability. It has been hypothesized that males' superior spatial ability is responsible for their superior quantitative ability, although "spatial ability" is a term that includes a number of different tasks. Men tend to outperform women in tasks that involve spatial visualization and mental rotation. Women do somewhat better (although not always as well) in comparison with men on tests of mathematics achievement. Moreover, as differences become smaller between males and females in their patterns of mathematics course-taking, the test differences appear to diminish.

Trend data show that although females may still have a small advantage in verbal ability, the gap has been eradicated for many tests. In the past few years there has been a decline in women's verbal scores and an increase in men's, especially in admission tests. At least one demographic analysis suggests that some of the trend data in admission test performance can be attributed to the changing nature of the test population. In quantitative ability there has been some convergence of scores, owing mainly to an improvement in women's scores. However, this convergence is not evident at the higher score ranges, where men are still disproportionately represented and the gap may even have increased. The remaining difference does not seem to be explained solely by differences in the courses taken by males and females, as some critics have suggested. And although the gap has been narrowing for both verbal and quantitative differences, the difference in test scores is still very pronounced with respect to quantitative tests. In general, for the population at large, there has been a decline in test scores, although the most recent National Assessment of Educational Progress has demonstrated some improvement in the lower score ranges and among minority groups.

The review also treats some of the leading hypotheses about the causes of gender differences in test performance and examines a selection of studies related to biological differences, sex-role development, and differences in social and educational phenomena, from interests and attitudes to patterns of course-taking. School experiences represent an important area of influence on test scores, and there is growing evidence that males and females have quite different educational histories and school experiences. A final area of research is related to characteristics of the tests themselves, particularly items that appear to function differently for males and females.

No single explanation captures all the variance in the differences between males and females in the quantitative domain. Patterns of course-taking, attitudes toward mathematics, differences in the achievement motivation of males and females, and some characteristics of the tests themselves may contribute to the differences but fail to explain all of them. The cumulative effects of early socialization patterns and different educational experiences have also been identified as likely influences on the performance differences, but these effects are difficult to assess with precision and require longitudinal data. Some researchers continue to search for biological antecedents, particularly of the differences in spatial ability, and of the special skills of the mathematically precocious, a disproportionate number of whom are male. Because there is a likelihood that all these possible causes are involved to some extent in the genesis and maintenance of gender differences in test performance, the most promising approaches to explanation of the phenomena are probably multivariate.

Continued research is needed on a number of fronts, to continue to document the trends in performance among subgroups of the total population; to examine the correlates and antecedents of disparities in performance among subgroups; and to analyze the types of tasks and items that evoke the largest differences between groups.

DISCUSSION

The topic of sex differences, particularly as they are implicated in intellectual performance, is, as Chipman (1988) points out in a review of Hyde and Linn (1986), a sexy topic. The media attention afforded the subject may serve more to polarize opinion about such differences than to foster understanding of them.

Insofar as there is convergence among studies

(and there continues to be controversy about many of the major findings), it is that the disparities between the sexes are slowly (perhaps too slowly for some) diminishing. Males appear to have caught up with females in tests of verbal ability and achievement, to the point where the absolute differences can be considered insignificant. And females have gained on but not equaled males in performance on some tests of mathematics ability and achievement, accompanied by increases in their participation in mathematics and their interest and self-confidence in that domain. With the exception of some limited domains of spatial ability, and performance at the top levels of mathematics achievement, women are improving their position relative to men. In nationally representative samples, against a backdrop of declining performance through the early 1980s, the tendency for the disparities to have diminished is quite evident.

Why, then, the continued concern? There are, perhaps, two main reasons. The first has to do with the social consequences of even the smallest of differences where large numbers of individuals are concerned. The second has to do with the ways in which test performance may affect subsequent motivation, attitudes, and behavior.

There are real, quantifiable educational and social consequences of test performance. Admission to college and to postsecondary educational programs, eligibility for special programs and for scholarships are often based on performance on the tests that have been reviewed here. Qualification for certain careers also depends in many cases on measures of the sort that show the largest differences between males and females. Even small differences can add up to major effects in the aggregate. Slight shifts in the ratio of male to female superiority in a domain can alter the nature of the population that qualifies for special awards, scholarships, programs, and educational opportunities. The recent reversal of advantage in the verbal domain on some measures from females to males, combined with the continuing disadvantage of females on quantitative measures, will undoubtedly exercise a substantial negative effect on the numbers of females that qualify for such awards, scholarships, and opportunities.

The second concern, which is related, is less easily quantified. Many of the studies reviewed here suggest that males and females may be affected differently by their success (or lack of success) as that success is reflected in test performance. Lesser performance on (say) measures of mathematical skill, whatever their origin, may cause females to lower their aspirations, lose their self-confidence, take courses in areas other than the quantitative ones, and/or conclude that certain domains are the province of males. The most visible form of the concern is the attention given the relative shortages of women in science and mathematics (e.g., Chipman et al. 1985; Fox et al. 1980). One product of this concern is a search for intervention strategies that can break into the cycle to increase the choices for females.

Both of these concerns demand that research into the nature and causes of sex differences in test performance continue. It is important to continue to examine the social correlates of the differences, both for reasons of increasing our understanding of the phenomenon and to inform efforts at intervention. For such studies, large-scale databases and multivariate methodologies are probably the most productive approaches. It is equally important to understand the cognitive processes that underlie the differences. In this respect, item-level studies and protocol analysis in the context of experimental studies are useful tools for continued research. There is a need for more studies that control for some of the variables that are confounded in research with existing test populations. This suggests some experimental studies with specially formulated items set in tests that vary with respect to item format, content, speededness, context for individual items, and cognitive requirements. Finally, although there is undoubtedly a need for studies that investigate possible biological or physiological correlates of gender differences in test performance, these seem less attractive intuitively because of their general intransigence and lack of potential for intervention.

There is an additional literature on gender differences that is not explored at all in this review, but which is exemplified in articles by Wittig (1985), Deaux (1985), and others. This literature is metatheoretical and asks questions about the basic premises underlying the study of gender. Some of the questions are relevant to the apparent contradictions that mark the gender-differences literature. Wittig, for example, mentions the tension between scholarship and advocacy that is present in psychology generally, but that is a major problem in the psychology of gender (as it is in the psychology of race). Disagreements about the significance of effect sizes may well be rooted in that tension. A related issue concerns the tension between scientific and humanistic values. The search for biological (as opposed to social or educational) causes seems to reflect that tension. Although discussion of such issues is beyond the scope of this review, the issues are mentioned (and some references are provided) so that interested readers and researchers can examine a range of perspectives if they choose to do so.

REFERENCES

Altman, R. A., and Holland, P. W. 1977. *A summary of data collected from Graduate Record Examinations test-takers during 1975-76.* ETS Data Summary Report No. 1. Princeton, N.J.: Educational Testing Service.

American College Testing Program. 1988. *ACT Assessment Program technical manual.* Iowa City, Iowa: ACT.

Anastasi, A., ed. 1958. *Differential psychology,* 3rd ed. New York: Macmillan.

Annett, M. 1980. Sex differences in laterality: meaningfulness vs. reliability. *The Behavioral and Brain Sciences* 3:227-63.

Applebee, A. N., Langer, J. A., and Mullis, I. V. S. 1986a. *Writing: Trends across the decade, 1974-1984.* ETS Report No. 15-W-01. Princeton, N.J.: NAEP, Educational Testing Service.

Applebee, A. N., Langer, J. A., and Mullis, I. V. S. 1986b. *The writing report card: Writing achievement in American schools.* ETS Report No. 15-W-02. Princeton, N.J.: NAEP, Educational Testing Service.

Armstrong, J.M. 1981. Achievement and participation of women in mathematics: Results of two national surveys. *Journal of Research in Mathematics Education* 12(5):356-72.

Armstrong, J. M. 1985. A national assessment of participation and achievement of women in mathematics. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation,* pp. 59-94. Hillsdale, N.J.: Lawrence Erlbaum.

Ash, B. F. 1986. *Identifying learning styles and matching strategies for teaching and learning.* ERIC Document Reproduction Service No. ED 270 142.

Association of American Medical Colleges [AAMC]: Section for Student and Educational Programs. 1987. *Percentile rank ranges for MCAT areas of assessment: 1987 summary of score distributions.* Washington, D.C.: AAMC.

Becker, B. J., and Hedges, L. V. 1984. Meta-analysis of cognitive gender differences: A comment on an analysis by Rosenthal and Rubin. *Journal of Educational Psychology* 76(4):583-87.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. 1986. *Women's ways of knowing: The development of self, voice and mind.* New York: Basic Books.

Bem, S. L., and Bem, D. J. 1976. Training the woman to know her place: The power of nonconscious ideology. In S. Cox, ed., *Female psychology: The emerging self,* pp. 180-190. Chicago: Science Research Associates.

Benbow, C. P. 1986. Physiological correlates of extreme intellectual precocity. *Neuropsychologia* 24:719-25.

Benbow, C. P. 1988. Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences,* in press.

Benbow, C. P., and Benbow, R. M. 1984. Biological correlates of high mathematical reasoning ability. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds, *Progress in brain research, Vol. 61: Sex differences in the brain,* pp. 469-90. New York: Elsevier.

Benbow, C. P., and Stanley, J. C. 1980. Sex differences in mathematical ability: Fact or artifact? *Science* 210:1262-64.

Benbow, C. P., and Stanley, J. C. 1981. Mathematical ability: Is sex a factor? *Science* 212:118-19.

Benbow, C. P., and Stanley, J. C. 1982. Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal* 19(4):598-622.

Benbow, C. P., and Stanley, J. C. 1983a. Differential course-taking hypothesis revisited. *American Educational Research Journal* 20(4):469-573.

Benbow, C. P., and Stanley, J. C. 1983b. Sex differences in mathematical reasoning ability: More facts. *Science* 222:1029-31.

Benbow, C. P., Stanley, J. C., Zonderman, A. B., and Kirk, M. K. 1983. Structure of intelligence of intellectually precocious children and of their parents. *Intelligence* 7:129-52.

Ben-Chaim, D., Lappan, G., and Houang, R. T. 1988. The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal* 25(1):51-71.

Bleier, R. 1984. *Science and gender: A critique of biology and its theories on women.* New York: Pergamon.

Bleier, R. 1987. Science and belief: A polemic on sex differences research. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 111-30. Indianapolis: Indiana University Press.

Block, J. H. 1976. Issues, problems, and pitfalls in assessing sex differences: A critical review of *The Psychology of Sex Differences. Merrill-Palmer Quarterly* 22(4):283-308.

Block, J. H. 1983. Differential premises arising from differential socialization of the sexes: Some conjectures. *Child Development* 54:1335-54.

Boswell, S. L. 1985. The influence of sex-role stereotyping on women's attitudes and achievement in mathematics. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation,* pp. 175-197. Hillsdale, N.J.: Lawrence Erlbaum.

Breland, H. M. 1977. *Group comparisons for the Test of Standard Written English.* ETS RDR 77-78, No. 1, Research Bulletin No RB-77-15, Princeton, N.J.: Educational Testing Service.

Breland, H. M., and Griswold, P. A. 1982. Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology* 74(5):713-21.

Bridgeman, B. 1988. *Comparative validity of multiple-choice and free-response advanced placement biology items.* Research report draft, submitted for review. Princeton, N.J.: Educational Testing Service.

Brody, L. E. 1987. *Gender differences on standardized examinations used for selecting applicants to graduate and professional schools.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Brooks-Gunn, J., and Matthews, W. S. 1979. *He & she: How children develop their sex-role identity.* Englewood Cliffs, N.J.: Prentice-Hall.

Burton, N. W. 1987, April. *Trends in the verbal scores of women taking the SAT in comparison to trends in other voluntary testing programs.* Paper presented at the

annual meeting of the American Educational Research Association, Washington, D.C.

Burton, N. W. 1988, April. *Modeling women's performance on the SAT.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Burton, N. W., Lewis, C., and Robertson, N. 1988, April. Draft. *SAT gender differences controlled for population trends.* Princeton, N.J.: Educational Testing Service.

Butler, S. 1984. Sex differences in human cerebral function. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 443-55. New York: Elsevier.

Caplan, P. J., MacPherson, G. M., and Tobin, P. 1985. Do sex-related differences in spatial abilities exist? *American Psychologist* 40(7):786-99.

Carlton, S. T. 1987, July. *Differences in male and female performance on standardized verbal tests.* Paper presented at the Third International Interdisciplinary Congress on Women, Dublin, Ireland.

Cherry, L., and Lewis, M. 1975. Mothers and two-year-olds: A study of sex-differentiated aspects of vebal interaction. *Developmental Psychology* 12(4):278-82.

Chipman, S. F. 1988, March/April. Far too sexy a topic [Review of *The psychology of gender differences: Advances through meta-analysis*]. *Educational Researcher* 17:46-49.

Chipman, S. F. 1988, April. *Word problems: Where test bias creeps in.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Chipman, S. F., Brush, L. R., and Wilson, D. M. eds. 1985. *Women and mathematics: Balancing the equation.* Hillsdale, N.J.: Lawrence Erlbaum.

Christensen, C. 1988. Personal communication.

Clark, M.J. and Grandy, J. 1984. *Sex differences in the academic performance of Scholastic Aptitude Test takers.* College Board Report No. 84-8. New York: College Entrance Examination Board.

Cohen, J. 1977. *Statistical power analysis for the behavioral sciences.* Revised ed. New York: Academic Press.

College Entrance Examination Board, Admissions Testing Program. 1986. *National college-bound seniors, 1985.* Princeton, N.J.: Educational Testing Service.

College Entrance Examination Board, 1987. *College-bound seniors: 1987 profile of SAT and Achievement Test takers.* Princeton, N.J.: Educational Testing Service.

Condry, J., and Dyer, S. 1976. Fear of success: Attributions of cause to the victim. *Journal of Social Issues* 32:63-83.

Connor, J. M., and Serbin, L. A. 1985. Visual-spatial skill: Is it important for mathematics? Can it be taught? In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation,* pp. 151-74. Hillsdale, N.J.: Lawrence Erlbaum.

Cox, P. W., and Witkin, H. A. 1978. *Field dependence-independence and psychological differentiation: Bibliography with index, Supplement No. 3.* ETS Research Bulletin No. 78-8. Princeton, N.J.: Educational Testing Service.

Crosson, C. W. 1984. Age and field independence among women. *Experimental Aging Research,* 10:165-70.

Dauber, S. L. 1987, April. *Sex differences on the SAT-M, SAT-V, TSWE, and ACT among college-bound high school students.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Deaux, K. 1985. Sex and gender. *Annual Review of Psychology* 36:49-81.

Deaux, K., and Major, B. 1987. Putting gender into context: An interactive model of gender-related behavior. *Psychological Review* 94(3);369-89.

Department of Defense, Office of the Assistant Secretary of Defense. 1982. *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery,* Washington, D.C.: Department of Defense.

Diener, C. I., and Dweck, C. S. 1980. An analysis of learned helplessness: The processing of success. *Journal of Personality and Social Psychology* 39:940-50.

Dix, L. S. 1987. *Women: Their underrepresentation and career differentials in science and engineering.* Proceedings of a workshop. Washington, D.C.: National Academy Press, Office of Scientific and Engineering Personnel.

Donlon, T. F., ed. 1984. *The College Board Technical Handbook for the Scholastic Aptitide Test and Achievement Tests.* New York: College Entrance Examination Board.

Donlon, T. F., Ekstrom, R. B., and Lockheed, M. E. 1979. The consequences of test bias in the content of major achievement test batteries. *Measurement and Evaluation in Guidance.* 11(4):202-16.

Donlon, T. F., Hicks, M. M., and Wallmark, M. M. 1980. Sex differences in item responses on the Graduate Record Examination. *Applied Psychological Measurement* 4(1):9-20.

Doolittle, A. E. 1985, April. *Understanding differential item performance as a consequence of gender differences in academic background.* Paper presented at the annual meeting of the American Education Research Association, Chicago, Ill.

Doolittle, A. E. 1987, August. *Gender differences in performance on mathematics achievement items.* Paper presented at the annual meeting of the American Psychological Association, New York.

Doolittle, A. E., and Cleary, T. A. 1987. Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement* 24(2):157-66.

Dorans, N. J. 1982. *Technical review of SAT item fairness studies: 1975-1979.* ETS Statistical Report No. SR-82-9. Princeton, N.J.: Educational Testing Service.

Dorans, N. J., and Livingston, S. A. 1987. Male-female difference in SAT-Verbal ability among students of high SAT-Mathematical ability. *Journal of Educational Measurement* 24(1):65-71.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., and Chambers , D. L. 1988. *The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment.* Princeton, N.J.: The Nation's Report Card, NAEP, Educational Testing Service.

Dunn, B. R. 1988, April. *Gender differences in EEG patterns: Are they indexes of different cognitive strategies?* Paper presented at the annual meeting of the American

Educational Research Association, New Orleans, La.
Dweck, C. S., Davidson, W., Nelson, S., and Enna, B. 1978. Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental Psychology* 14(3):268–76.
Eccles, J. S. 1985. Sex differences in achievement patterns. In T. Sonderegger, ed., *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.
Eccles, J. S. 1986. Gender-roles and women's achievement. *Educational Researcher* 15:15–19.
Eccles, J. S. 1987. Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly* 11:135–72.
Eccles (Parsons), J. 1983. Expectancies, values, and academic behaviors. In J. T. Spence, ed. *Achievement and achievement motives: Psychological and sociological approaches.* San Francisco: Freeman.
Eccles (Parsons), J., Adler, T., and Meece, J. L.: 1984. Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology* 46(1):26–43.
Educational Testing Service. 1987. *A summary of data collected from Graduate Record Examinations test-takers during 1985-1986.* ETS Data Summary Report No. 11. Princeton, N.J.: Educational Testing Service.
Ekstrom, R., Goertz, M., and Rock, D. 1988. *Education and American youth.* London: Falmer Press.
Ethington, C. A., and Wolfle, L. M. 1986. A structural model of mathematics achievement for men and women. *American Educational Research Journal* 23(1):65–75.
Fagot, B. I. 1978. The influence of sex of child on parental reactions to toddler children. *Child Development* 49:459–65.
Farmer, H. S. 1987, March. A multivariate model for explaining gender differences in career and achievement motivation. *Educational Researcher,* 16:5–9.
Farr, R., Courtland, M. C., and Beck, M. D. 1984, December. Scholastic Aptitude Test performance and reading ability. *Journal of Reading,* 208–14.
Fausto-Sterling, A. 1985. *Myths of gender: Biological theories about women and men.* New York. Basic Books.
Feingold, A. 1988. Cognitive gender differences are disappearing. *American Psychologist* 43(2):95–103.
Fennema, E. 1974. Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education* 5:126–29.
Fennema, E., and Ayer, M. J., eds. 1984. *Women and education: Equity or equality?* Berkeley, Calif.: McCutchan.
Fennema, E., and Carpenter, T. 1981. The second National Assessment and sex-related differences in mathematics. *Mathematics Teacher* 74:554–59.
Fennema, E., and Tartre, L. A. 1985. The use of spatial visualization in mathematics by girls and boys. *Journal for Research in Mathematics Education* 16(3)184–206.
Fox, L. H., Brody, L., and Tobin, D., eds. 1980. *Women and the mathematical mystique.* Baltimore, Md.: Johns Hopkins University Press.
Fox, L. H., Fennema, E., and Sherman, J. 1977. *Women and mathematics: Research perspectives for change.* NIE Papers in Education and Work, No. 8. Washington, D.C.: National Institute of Education.
Freed, N. H. 1983. Foreseeably equivalent math skills of men and women. *Psychological Reports* 52:334.
Frieze, I. H. 1980. Beliefs about success and failure in the classroom. In J. McMillan, ed., *The social psychology of school learning.* New York: Academic Press.
Frieze, I. H., Whitley, B. E., Hanusa, B. H., and McHugh, M. 1982. Assessing the theoretical models for sex differences in casual attributions for success and failure. *Sex Roles* 8:333–45.
Gilligan, C. 1987. Remapping development: The power of divergent data. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 77–94. Indianapolis: Indiana University Press.
Golden, M. and Birns, B. 1983. Social class and infant intelligence. In M. Lewis, ed., *Origins of intelligence: Infancy and early childhood.* pp. 347–398. New York: Plenum Press.
Goodenough, D. R., and Witkin, H. A. 1977. *Origins of field-dependent and field-independent cognitive styles.* ETS Research Bulletin No. 77–9. ERIC Document Reproduction Service No. ED 150 155. Princeton, N.J.: Educational Testing Service.
Goodison, M. B. 1982. *A summary of data collected from Graduate Record Examinations test-takers during 1980-81.* ETS Data Summary Report No. 6. Princeton, N.J.: Educational Testing Service.
Grandy, J. 1987, October. *Trends in the selection of science, mathematics, or engineering as major fields of study among top-scoring SAT takers.* ETS Research Report No. 87–39. Princeton, N.J.: Educational Testing Service.
Grandy, J. 1987. October. *Ten-year trends in SAT scores and other characteristics of high school seniors taking the SAT and planning to study mathematics, science, or engineering.* ETS Research Report No. 87–49. Princeton, N.J.: Educational Testing Service.
Grandy, J., and Courtney, R. 1985. *Factors contributing to the changing characteristics of prospective humanities majors: 1975-1984.* Grant No. OP-20193–84. Princeton, N.J.: Educational Testing Service.
Grant, C. A., and Sleeter, C. E. 1986. Race, class, and gender in education research: An argument for integrative analysis. *Review of Educational Research* 56(2):195–211.
Grieb, A., and Easley, J. 1984. A primary school impediment to mathematics equity: Case studies in rule-dependent socialization. In M. W. Steinkamp and M. L. Maehr, eds., *Advances in motivation and achievement. Vol. 2: Women in science.* Greenwich, Conn.: Jai Press.
Haertel, G. D., Walberg, H. J., Junker, L., and Pascarella, E. T. 1981. Early adolescent sex differences in science learning: Evidence from the National Assessment of Educational Progress. *American Educational Research Journal* 18(3):329–41.
Halpern, D. F. 1986. *Sex differences in cognitive abilities.* Hillsdale, N.J.: Lawrence Erlbaum.
Harter, S. 1983. A model of intrinsic mastery motivation in children: Individual differences and developmental change. Minnesota Symposium on Child Development 14. Hillsdale, N.J.: Lawrence Erlbaum Associates.
Heister, G. 1984. Sex differences in visual half-field superiority as a function of responding hand and motor demands. In G. J. De Vries, J. P. C. De Bruin, H. B. M.

Uylings and M. A. Corner, eds., *Progress in brain*
Horner, M. S. 1970. Femininity and successful achievement: A basic inconsistency. In J. M. Bardwick, et al., eds., *Feminine personality and conflict.* Monterey, Calif.: Brooks/Cole.
Houston, A. C. 1983. Sex-typing. In P. H. Mussen, ed., *Handbook of child psychology,* vol. 4. New York: Wiley.
Huber, G. L. 1988, April. *Preference for learning situations and uncertainty orientation: A cross-cultural comparison.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Hudson, L. 1986. Item-level analysis of sex differences in mathematics achievement test performance. *Dissertation Abstracts International* 47(2). Order no. DA8607283.
Hyde, J. S., Geiringer, E. R., and Yen, W. M. 1975. On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. *Multivariate Behavioral Research* 10:289–309.
Hyde, J. S. 1981. How large are cognitive gender differences? A meta-analysis using w^2 and d. *American Psychologist* 36(8):892–901.
Hyde, J. S., and Linn, M. C., eds. 1988. *The psychology of gender: Advances through meta-analysis.* Baltimore, Md.: Johns Hopkins University Press.
Hyde, J. S., and Linn, M. C. 1988, in press. A meta-analysis of gender differences in verbal abilities. *Psychological Bulletin.*
Jacklin, C. N. 1987. Feminist research and psychology. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 94–107. Indianapolis: Indiana University Press.
Jacobs, J. E. 1978. *Perspectives on women and mathematics.* Columbus, Ohio: ERIC Clearinghouse for Science, Mathematics and Environmental Education.
Jacobs, J. E., and Eccles, J. S. 1985, March. Gender differences in math ability: The impact of media reports on parents. *Educational Researcher,* 14:20–25.
Jones, L. V., Davenport, E. C., Bryson, A., Bekhuis, T., and Zwick, R. 1986. Mathematics and science test scores as related to courses taken in high school and other factors. *Journal of Educational Measurement* 23(3):197–208.
Jones, R. F. 1984. *Women and the MCAT: An overview of research in progress.* Paper presented at the annual meeting of the Association of American Medical Colleges, Chicago, Ill.
Jones, R. F. and Vanyur S. 1985, April. *An investigation of gender-related test bias for the Medical College Admission Test.* Paper presented at the meeting of the National Council on Measurement in Education, Chicago, Ill.
Kagan, J. 1964. The acquisition and significance of sex-*research, Vol. 61: Sex differences in the brain,* pp. 457–468. New York: Elsevier.
Hilton, T. L., and Berglund, G. W. 1974. Sex differences in mathematics achievement—a longitudinal study. *The Journal of Educational Research,* 67:231–37.
Hoffman, L. W. 1977. Changes in family roles, socialization and sex differences. *American Psychologist* 32:644–57.
Hogrebe, M. C., Nist, S. L., and Newman, I. 1985. Are there gender differences in reading achievement? An investigation using the High School & Beyond data. *Journal of Educational Psychology* 77(6):716–24.
typing and sex-role identity. In M. Hoffman and L. Hoffman, eds., *Review of child development research,* vol. 1. New York: Russell Sage.
Kahle, J. B. 1984. *Girls in school/women in science: A Synopsis.* Paper presented at the annual Women's Studies Conference, Greeley, Colo. ERIC Document Reproduction Service No. ED 243 785.
Kahle, J. B. and Lakes, M. K. 1983. The myth of equality in science classrooms. *Journal of Research in Science Teaching* 20(2):131–40.
Karmos, A. H., and Karmos, J. S. 1984, July. Attitudes toward standardized achievement tests and their relation to achievement test performance. *Measurement and Evaluation in Counseling and Development,* 56–66.
Kavrell, S. M., and Peterson, A. C. 1984. Patterns of achievement in early adolescence. In M. W. Steinkamp and M. L. Maehr, eds., *Advances in motivation and achievement. Vol. 2: Women in science.* Greenwich, Conn.: Jai Press.
Keller, E. F. 1985. *Reflections on gender and science.* New Haven, Conn.: Yale University Press.
Kimura, D., and Harshman, R. A. 1984. Sex differences in brain organization for verbal and non-verbal functions. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 423–41. New York: Elsevier.
Kirsch, I. S., and Jungeblut, A. 1986. *Literacy: Profiles of America's young adults.* Princeton, N.J.: National Assessment of Educational Progress (NAEP), Educational Testing Service.
Klein, S. S., ed. 1980. *Sex equity in education: NIE-sponsored projects and publications.* Washington, D.C.: National Institute of Education.
Klein, S. S., ed. 1985. *Handbook for achieving sex equity through education.* Baltimore, Md.: Johns Hopkins University Press.
Laing, J., Engen, H., and Maxey, J. 1987. *Relationships between ACT test scores and high school courses.* Research report. Iowa City, Iowa: American College Testing Program.
Law School Admissions Council. 1988. *LSAC/LSAS National Statistical Report 1982–83 through 1986–87.* Newtown, Pa.: LSAC.
Lawrence, I. M., Curley, W. E., and McHale, F. J. 1987. *Differential item functioning of SAT-Verbal reading subscore items for male and female examinees.* ETS Research Report, in press. Princeton, N.J.: Educational Testing Service.
Lehrke, R. G. 1974. *X-linked mental retardation and verbal disability.* New York: Intercontinental Medical Book.
Leinhardt, G., Seewald, A. M., and Engel, M. 1979. Learning what's taught: Sex differences in instruction. *Journal of Educational Psychology* 71(4):432–39.
Lenny, E. 1977. Women's self-confidence in achievement settings. *Psychological Bulletin* 84:1–13.
Lepper, M. 1985. Microcomputers in education: Motivational and social issues. *American Psychologist* 40:1–18.
Levine, D. U., and Ornstein, A. C. 1983. Sex differences in ability and achievement. *Journal of Research and Development in Education* 16(2):66–72.
Lewis, M., and Freedle, R. 1973. The mother-infant dyad.

In P. Pliner, L. Kranes, and T. Alloway, eds., *Communication and affect: Language and thought.* New York: Academic Press.

Licht, B. G., and Dweck, C. S. 1983. Sex differences in achievement orientations: Consequences for academic choices and attainments. In M. Marland, ed., *Sex differentiation and schooling.* London: Heinemann.

Linn, R. L. 1982. Ability testing: Individual differences, predictions and differential prediction. In A. Wigdor and W. Garner, eds., *Ability testing: Uses, consequences and controversies,* pp. 335–38. Washington, D.C.: National Academy Press.

Linn, M. C. 1988, May. *Trends in the magnitude and nature of cognitive gender differences: Implications for the SAT.* Paper presented at ETS Seminar, Princeton, N. J.

Linn, M. C., De Benedictis, T., Delucchi, K., Harris, A., and Stage, E. 1987. Gender differences in National Assessment of Educational Progress science items: What does "I don't know" really mean? *Journal of Research in Science Teaching* 24(3):267–78.

Linn, M. C., and Hyde, J. S. 1986, April. *Gender differences in verbal ability: A meta-analysis.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Linn, M. C., and Peterson, A. C. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development* 56:1479–98.

Linn, M. C., and Peterson, A. C. 1986. A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde and M. C. Linn, eds., *The Psychology of gender: Advances through meta-analysis,* pp. 67–101. Baltimore, Md.: Johns Hopkins University Press.

Lockheed, M. E. 1984. Sex segregation and male preeminence in elementary classrooms. In E. Fennema and M. J. Ayer, eds., *Women and education: Equity or equality?* pp. 117–35. Berkeley, Calif.: McCutchan.

Lockheed, M. E., Thorpe, M., Brooks-Gunn, J., Casserly, P., and McAloon, A. 1985. *Sex and ethnic differences in middle school mathematics, science and computer science: What do we know?* A report submitted to The Ford Foundation. Princeton, N.J.: Educational Testing Service.

Loewen, J. W., Rosser, P., and Katzman, J. 1988, April. *Gender bias in SAT items.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Lubetkin, J. 1988, April. The Scholastic Aptitude Test: A valid and unbiased predictor of college performance? Unpublished B.A. Thesis, Princeton University, Princeton, N.J.

Lupkowski, A. E. 1987, April. *Sex differences on the Differential Aptitude Test.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

McClelland, D. C. 1961. *The achieving society.* Princeton, N. J.: Van Nostrand.

Maccoby, E. E., 1966. *The development of sex differences.* Stanford, Calif.: Stanford University Press.

Maccoby, E. E., and Jacklin, C. N. 1974. *The Psychology of sex differences.* Stanford, Calif.: Stanford University Press.

McGoneghy, J. I. 1987, April. *Mathematics attitudes and achievement: Gender differences in a multivariate context.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C. ERIC Document Reproduction Service No. ED 284 742.

McGlone, J. 1980. Sex differences in human brain asymmetry: A critical survey. *The Behavioral and Brain Sciences* 3:215–63.

McPeek, W. M., and Wild, C. L. 1987, August. *Characteristics of quantitative items that function differently for men and women.* Paper presented at the annual meeting of the American Psychological Association, New York.

McPeek, W. M., and Wild, C. L. 1987, April. *Identifying differentially functioning items in the NTE core battery.* Unpublished research report. Princeton, N.J.: Educational Testing Service.

Meehan, A. M. 1984. A meta-analysis of sex differences in formal operational thought. *Child Development* 55:1110–24.

Messick, S. 1976. Personality consistencies in cognition and creativity. In S. Messick, ed., *Individuality in learning: Implications of cognitive style and creativity for human development,* pp. 4–22. San Francisco: Jossey-Bass.

Messick, S. 1984. The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist* 19(2):59–74.

Milton, G. A. 1957. The effects of sex-role identification upon problem-solving skill. *Journal of Abnormal and Social Psychology* 55:208–12.

Mullis, I. V. S. 1987, April. *Trends in performance for women taking the NAEP reading and writing assessment.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Murphy, R. J. L. 1982. Sex differences in objective test performance. *British Journal of Educational Psychology* 52:213–19.

National Assessment of Educational Progress. 1983. *The third national mathematics assessments: Results, trends and issues.* Denver, Colo.: Education Commission of the States.

National Assessment of Educational Progress. 1985. *The reading report card. Progress toward excellence in our schools: Trends in reading over four national assessments, 1971–1984.* ETS Report No. 15-R-01. Princeton, N.J.: Educational Testing Service.

National Assessment of Educational Progress. 1986. NAEP 1986 mathematics assessment, weighted W.A.R.M. background factor percentages and mean math proficiency composites. Unpublished raw data.

Newcombe, N., and Dubas, J. S. 1987. Individual differences in cognitive ability: Are they related to timing of puberty? In R. M. Lerner and T. L. Foche, ed., *Biological-psychosocial interactions in early adolescence.* pp. 249–302. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Nisbett, R. E., and Wilson, T. D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84:231–59.

Noble, J., and McNabb. T. 1988, April. *Differential coursework in high school: Implications for performance on the ACT assessment.* Paper presented at the annual meeting of the American Educational Research Asso-

ciation, New Orleans, La.

Nyborg, H. 1984. Performance and intelligence in hormonally different groups. In G. J. DeVries, et al., eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 491–508. New York: Elsevier.

Paley, V. G. 1984. *Boys and girls: Superheroes in the doll corner.* Chicago, Ill.: University of Chicago Press.

Pallas, A. M., and Alexander, K. L. 1983. Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. *American Educational Research Journal* 20(2):165–82.

Paulhus, D., and Schaeffer, D. R. 1981. Sex differences in the impact of older and number of younger siblings on scholastic aptitude. *Social Psychology Quarterly* 44:363–68.

Pearlman, M. A. 1987, April. *Trends in women's total score and item performance on verbal measures: Five forms of the GRE: Verbal items that display large Mantel-Haenszel DIF values.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Pennock-Roman, M., Rock, D. A., and Enright, M. K. 1988, January. *Language background and test validity for Hispanic-American students. Part I: Comparisons between Hispanic and non-Hispanic-White groups.* Unpublished manuscript. Princeton, N.J.: Educational Testing Service.

Peterson, A. C. 1983. Pubertal change and cognition. In J. Brooks-Gunn and A. C. Peterson, eds., *Girls at puberty: Biological and psychological perspectives,* pp. 179–98. New York: Plenum Press.

Peterson, P. L. and Fennema, E. 1985. Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal* 22(3):309–35.

Ramist, L. 1984. Predictive validity of the ATP tests. In T. Donlon, ed., *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests.* New York: The College Board.

Ramist, L., and Arbeiter, S. 1986. *Profiles, college-bound seniors, 1985.* New York: College Entrance Examination Board.

Raymond, C. L., and Benbow, C. P. 1986. Gender differences in mathematics: A function of parental support and student sex typing? *Development Psychology* 22(6):808–19.

Rheingold, H. L., and Cook, K. V. 1975. The content of boys' and girls' rooms as an index of parent behavior. *Child Development* 46:459–63.

Rock, D. A., Goertz, M. E., Ekstrom, R. B., Hilton, T. L., and Pollack, J. 1984, December. *Factors associated with test score decline.* Briefing paper. Princeton, N.J.: Educational Testing Service.

Rock, D. A., Hilton, T. L., Pollack, J., Ekstrom, R. B., and Goertz, M. E. 1985. *Psychometric analysis of the NLS and the High School and Beyond Test Batteries.* NCES Report No. 85-218. Washington, D.C.: National Center for Education Statistics.

Rosenthal, R., and Rubin, D. B. 1982. Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology* 74(5):708–12.

Rutter, M. 1977. Individual differences. In M. Rutter and L. Hersov, eds., *Child Psychiatry: Modern Approaches.* Oxford, England: Blackwell Scientific.

Sadker, M., and Sadker, D. 1985b, March. Sexism in the schoolroom of the 80's. *Psychology Today* 54–57.

Sadker, M., and Sadker, D. 1985a, January. Is the O.K. classroom O.K.? *Phi Delta Kappan.*

Sadker, M., and Sadker, D. 1986, March. Sexism in the classroom: From grade school to graduate school. *Phi Beta Kappan.*

Schickedanz, J. A. 1973. The relationship of sex-typing of reading to reading achievement and reading choice behavior in elementary school boys. *Dissertation Abstracts* 34(12A Pt. 1):7645.

Senk, S., and Usiskin, Z. 1983. Geometry proof writing: A new view of sex differences in mathematics ability. *American Journal of Education* 91:187–201.

Sherman, J. A. 1967. Problems of sex differences in space perception and aspects of intellectual functioning. *Psychological Review* 74:290–99.

Sherman, J. A. 1978. *Sex-related cognitive differences: An essay on theory and evidence.* Springfield, Ill.: Charles C. Thomas.

Slavin, R. E. 1978. Effects of student teams and peer tutoring on academic achievement and time-on-task. *Journal of Experimental Education* 48:252–57.

Stafford, R. E. 1972. Hereditary and environmental components of quantitative reasoning. *Review of Educational Research* 42:183–201.

Stallings, S. J. 1985. School, classroom and home influences on women's decisions to enroll in advanced mathematics courses. In S.F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation.* Hillsdale, N.J.: Lawrence Erlbaum.

Stanley, J. C. 1982, March. Identification of intellectual talent. In W. B. Schrader, ed., *New directions for testing and measurement: Measurement, guidance, and program improvement,* no. 13. San Francisco: Jossey-Bass.

Stanley, J. 1987, April. *Sex differences on the College Board Achievement Tests and the Advanced Placement Examinations.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Steelman, L. C., and Marcy, J. A. 1983. Sex differences in the impact of the number of older and younger siblings on IQ performance. *Social Psychology Quarterly* 46(2):157–62.

Steinkamp, M. W., and Maehr, M. L. 1984a. Gender differences in motivational orientations toward achievement in school science: A quantitative synthesis. *American Educational Research Journal* 21(1):39–59.

Steinkamp, M. W., and Maehr, M. L., eds. 1984b. *Advances in motivation and achievement. Vol. 2: Women in science.* Greenwich, Conn.: Jai Press.

Stockard, J., Schmuck, P. A., Kempner, K., Williams, P., Edson, S. K., and Smith, M. A. 1980. *Sex equity in education.* New York: Academic Press.

Swinton, S. S. 1987. *The predictive validity of the restructured GRE with particular attention to older students.* GRE Board Professional Report No. 83-25P, ETS RR No. 87-22. Princeton, N.J.: Educational Testing Service.

Tittle, C. K. 1986. Gender research and education. *American Psychologist* 41(10):1161–68.

Tobias, S. 1978. *Overcoming math anxiety.* New York: Norton.

Tyler, L. E. 1965. *The psychology of human differences.* 3rd ed. New York: Appleton-Century-Crofts.

Vandenberg, S. G. 1968. Primary mental abilities or general intelligence? Evidence from twin studies. In J. M. Thoday and A. S. Parkes, eds., *Genetic and environmental influences on behavior,* pp. 146–60. New York: Plenum.

Waber, D. P., Mann, M. B., Merola, J., and Moylan, P. M. 1985. Physical maturation rate and cognitive performance in early adolescence: A longitudinal examination. *Developmental Psychology* 21(4):666–81.

Weiner, B. 1979. A theory of motivation for some classroom experiences. *Journal of Educational Psychology* 71:3–25.

Weiner, B., Frieze, I. H., Kukla, A., Reed, L., Rest, S., and Rosenbaum, R. M. 1971. *Perceiving the causes of success and failure.* Morristown, N.J.: General Learning Press.

Welch, C. J., and Doolittle, A. E. 1988, April. *Gender-based differential item performance in English usage items.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Wendler, C. L. W., and Carlton, S. T. 1987, April. *An examination of SAT verbal items for differential performance by women and men: An exploratory study.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Wharton, Y. L. 1977. *List of hypotheses advanced to explain the SAT decline.* New York: College Entrance Examination Board.

Wheeler, P., and Harris, A. 1981. *Comparison of male and female performance on the ATP Physics Test.* CB Report No. 81–4. Princeton, N.J.: Educational Testing Service.

Whitley, B. E., Jr., and Frieze, I. H. 1985. *The effect of question wording style and research context on attributions for success and failure: A meta-analysis.* Paper presented at the annual meeting of the Eastern Psychological Association, Boston.

Wild, C. L. 1981. *A summary of data collected from Graduate Record Examinations test-takers during 1979–80.* ETS Data Summary Report No. 5. Princeton, N.J.: Educational Testing Service.

Wild, C. L., and Dwyer, C. A. 1980. Sex bias in selection. In L. J. Th. van der Kamp, W. F. Langerak, and D. N. M. de Gruijter, eds., *Psychometrics for educational debates,* pp. 153–68. New York: Wiley.

Wild, C. L., and McPeek, W. M. 1986, August. *Performance of the Mantel-Haenszel Statistic in identifying differentially functioning items.* Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Wilder, G., Casserly, P., and Burton, N. 1988. *Young SAT-takers: Two surveys.* College Board Report No. 88–1. New York: College Entrance Examination Board.

Wilkinson, L. C., and Marrett, C. B., eds. 1985. *Gender influences in classroom interaction.* New York: Academic Press.

Wise, L. L. 1985. Project TALENT: Mathematics course participation in the 1960s and its career consequences. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation,* pp. 25–58. Hillsdale, N.J.: Lawrence Erlbaum.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. G., and Karp, S. A. 1962. *Psychological differentiation.* New York: Wiley.

Witkin, H. A., and Goodenough, D. R. 1981. Cognitive styles: Essence and origins. *Psychological issues.* Monograph 51. New York: International Universities Press.

Witkin, H. A., Goodenough, D. R., and Oltman, P. K. 1979. Psychological differentiation: Current status. *Journal of Personality and Social Psychology* 37(7):1127–45.

Wittig, M. A. 1985. Metatheoretical dilemmas in the psychology of gender. *American Psychologist* 40(7):800–811.

Wittig, M. A., and Peterson, A. C., eds. 1979. *Sex-related differences in cognitive functioning: Developmental issues.* New York: Academic Press.

Wittig, M. A., Sasse, S. H., and Giacomi, J. 1984. Predictive validity of five cognitive skills tests among women receiving engineering training. *Journal of Research in Science Teaching* 21(5):537–46.

Wolleat, P. L., Pedro, J. D., Becker, A. D., and Fennema, E. 1980. Sex differences in high school students' casual attributions of performance in mathematics. *Journal for Research in Mathematics Education* 11:356–66.

Women on Words and Images. 1972. *Dick and Jane as victims: Sex stereotyping in children's readers.* Princeton, N.J.: Women on Words and Images.

Zajonc, R. B. 1986. The decline and rise of scholastic aptitude scores: A prediction derived from the confluence model. *American Psychologist* 41(8):862–67.

Zajonc, R. B., and Bargh, J. 1980a. Birth order, family size and decline of SAT scores. *American Psychologist* 35:662–68.

Zajonc, R. B., and Bargh, J. 1980b. The confluence model: Parameter estimation for six divergent data sets on family factors and intelligence. *Intelligence* 4:349–61.

Zerega, M. E., Haertel, G. D., Tsai, S.-L., and Walberg, H. J. 1986. Late adolescent sex differences in science learning. *Science Education* 70(4):447–60.

Zill, N. 1985. *Happy, healthy and insecure: A portrait of middle childhood in the United States.* New York: Cambridge University Press.

APPENDIX A. REFERENCES ARRANGED BY FORMAT AND TOPIC

Books

Anastasi, A., ed. 1958. *Differential psychology.* 3rd ed. New York: Macmillan.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. 1986. *Women's ways of knowing: The development of self, voice and mind.* New York: Basic Books.

Bleier, R. 1984. *Science and gender: A critique of biology and its theories on women.* New York: Pergamon.

Brooks-Gunn, J., and Matthews, W. S. 1979. *He & she: How children develop their sex-role identity.* Englewood Cliffs, N.J.: Prentice-Hall.

Chipman, S. F., Brush, L. R., and Wilson, D. M., eds. 1985. *Women and mathematics: Balancing the equation.* Hillsdale, N.J.: Lawrence Erlbaum.

Ekstrom, R., Goertz, M., and Rock, D. 1988. *Education and American youth.* London: Falmer Press.

Fausto-Sterling, A. 1985. *Myths of gender: Biological theories about women and men.* New York: Basic Books.

Fennema, E., and Ayer, M. J., eds. 1984. *Women and education: Equity or equality?* Berkeley, Calif.: McCutchan.

Fox, L. H., Brody, L., and Tobin, D., eds. 1980. *Women and the mathematical mystique.* Baltimore, Md.: Johns Hopkins University Press.

Fox, L. H., Fennema, E., and Sherman, J. 1977. *Women and mathematics: Research perspectives for change.* NIE Papers in Education and Work, No. 8. Washington, D.C.: National Institute of Education.

Halpern, D. F. 1986. *Sex differences in cognitive abilities.* Hillsdale, N.J.: Lawrence Erlbaum.

Hyde, J. S., and Linn, M. C., eds. 1986. *The psychology of gender: Advances through meta-analysis.* Baltimore, Md.: Johns Hopkins University Press.

Jacobs, J. E. 1978. *Perspectives on women and mathematics.* Columbus, Ohio: ERIC Clearinghouse for Science, Mathematics and Environmental Education.

Keller, E. F. 1985. *Reflections on gender and science.* New Haven, Conn.: Yale University Press.

Klein, S. S., ed. 1980. *Sex equity in education: NIE-sponsored projects and publications.* Washington, D.C.: National Institute of Education.

Klein, S. S., ed. 1985. *Handbook for achieving sex equity through education.* Baltimore, Md.: Johns Hopkins University Press.

McClelland, D. C. 1961. *The achieving society.* Princeton, N.J.: Van Nostrand.

Maccoby, E. E., ed. 1966. *The Development of sex differences.* Stanford, Calif.: Stanford University Press.

Maccoby, E. E., and Jacklin, C. N. 1974. *The Psychology of sex differences.* Stanford, Calif.: Stanford University Press.

Paley, V. G. 1984. *Boys and girls: Superheroes in the doll corner.* Chicago, Ill.: University of Chicago Press.

Sherman, J. A. 1978. *Sex-related cognitive differences: An essay on theory and evidence.* Springfield, Ill.: Charles C. Thomas.

Steinkamp, M. W., and Maehr, M. L., eds. 1984. *Advances in motivation and achievement. Vol. 2: Women in Science.* Greenwich, Conn.: Jai Press.

Stockard, J., Schmuck, P. A., Kempner, K., Williams, P., Edson, S. K. and Smith, M. A. 1980. *Sex equity in education.* New York: Academic Press.

Tobias, S. 1978. *Overcoming math anxiety.* New York: Norton.

Tyler, L. E. 1965. *The psychology of human differences.* 3rd ed. New York: Appleton-Century-Crofts.

Wilkinson, L. C., and Marrett, C. B., eds. 1985. *Gender influences in classroom interaction.* New York: Academic Press.

Wittig, M. A., and Petersen, A. C., eds. 1979. *Sex-related differences in cognitive functioning: Developmental issues.* New York: Academic Press.

Zill, N. 1985. *Happy, healthy and insecure: A portrait of middle childhood in the United States.* New York: Cambridge University Press.

Descriptive Data and Summary Reports

Altman, R. A., and Holland, P. W. 1977. *A summary of data collected from Graduate Record Examinations test-takers during 1975–76.* ETS Data Summary Report No. 1. Princeton, N.J.: Educational Testing Service.

Applebee, A. N., Langer, J. A., and Mullis, I. V. S. 1986a. *Writing: Trends across the decade, 1974–84.* ETS Report No. 15-W-01. Princeton, N.J.: NAEP, Educational Testing Service.

Applebee, A. N., Langer, J. A., and Mullis, I. V. S. 1986b. *The writing report card: Writing achievement in American schools.* ETS Report No. 15-W-02. Princeton, N.J.: NAEP, Educational Testing Service.

Armstrong, J. M. 1981. Achievement and participation of women in mathematics: Results of two national surveys. *Journal of Research in Mathematics Education* 12(5):356–72.

Association of American Medical Colleges [AAMC]: Section for Student and Educational Programs. 1987. *Percentile rank ranges for MCAT areas of assessment: 1987 summary of score distributions.* Washington, D.C.: AAMC.

Benbow, C. P., and Stanley, J. C. 1980. Sex differences in mathematical ability: Fact or artifact?: *Science* 210:1262–64.

Benbow, C. P., and Stanley, J. C. 1981. Mathematical ability: Is sex a factor? *Science* 212:118–19.

Benbow, C. P., and Stanley, J. C. 1982. Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal* 19(4):598–622.

Benbow, C. P., and Stanley, J. C. 1983a. Differential course-taking hypothesis revisited. *American Educational Research Journal* 20(4):469–573.

Benbow, C. P., and Stanley, J. C. 1983b. Sex differences in mathematical reasoning ability: More facts. *Science* 222:1029–31.

Benbow, C. P., Stanley, J. C., Zonderman, A. B., and Kirk, M. K. 1983, Structure of intelligence of intellectually precocious children and of their parents. *Intelligence* 7:129–52.

Breland, H. M. 1977. *Group comparisons for the Test of Standard Written English.* ETS RDR 77-78, No. 1, Research Bulletin No. RB-77-15. Princeton, N.J.: Educational Testing Service.

Brody, L. E. 1987. *Gender differences on standardized examinations used for selecting applicants to graduate and professional schools.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Burton, N. W. 1987, April. *Trends in the verbal scores of women taking the SAT in comparison to trends in other voluntary testing programs.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Christensen, C. 1988. Personal communication.

Clark, M. J., and Grandy, J. 1984. *Sex differences in the academic performance of Scholastic Aptitude Test takers.* College Board Report No. 84-8. New York: College Entrance Examination Board.

College Entrance Examination Board, Admissions Testing Program. 1986. *National college-bound seniors, 1985.* Princeton, N.J.: Educational Testing Service.

College Entrance Examination Board. 1987. *College-bound seniors: 1987 profile of SAT and Achievement Test takers.* Princeton, N.J.: Educational Testing Service.

Dauber, S. L. 1987, April. *Sex differences on the SAT-M, SAT-V, TSWE, and ACT among college-bound high school students.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Department of Defense, Office of the Assistant Secretary of Defense. 1982. *Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery.* Washington, D.C.: Department of Defense.

Dorans, N. J., and Livingston, S. A. 1987. Male-female difference in SAT-Verbal ability among students of high SAT-Mathematical ability. *Journal of Educational Measurement* 24(1):65-71.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., and Chambers, D. L. 1988. *The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment.* Princeton, N.J.: The Nation's Report Card, NAEP, Educational Testing Service.

Educational Testing Service. 1987. *A summary of data collected from Graduate Record Examinations test-takers during 1985-86.* ETS Data Summary Report No. 11. Princeton, N.J.: Educational Testing Service.

Farr, R., Courtland, M. C., and Beck, M. D. 1984, December. Scholastic Aptitude Test performance and reading ability. *Journal of Reading,* 208-14.

Feingold, A. 1988. Cognitive gender differences are disappearing. *American Psychologist* 43(2):95-103.

Fennema, E., and Carpenter, T. 1981. The second National Assessment and sex-related differences in mathematics. *Mathematics Teacher* 74:554-59.

Freed, N. H. 1983. Foreseeably equivalent math skills of men and women. *Psychological Reports* 52:334.

Goodison, M. B. 1982. *A summary of data collected from Graduate Record Examinations test-takers during 1980-81.* ETS Data Summary Report No. 6. Princeton, N.J.: Educational Testing Service.

Hilton, T. L., and Berglund, G. W. 1974. Sex differences in mathematics achievement—a longitudinal study. *The Journal of Educational Research* 67:231-37.

Hogrebe, M. C., Nist, S. L., and Newman, I. 1985. Are there gender differences in reading achievement? An investigation using the High School & Beyond data. *Journal of Educational Psychology* 77(6):716-24.

Jones, R. F. 1984. *Women and the MCAT: An overview of research in progress.* Paper presented at the annual meeting of the Association of American Medical Colleges, Chicago, Ill.

Kirsh, I. S., and Jungeblut, A. 1986. *Literacy: Profiles of America's young adults.* Princeton, N.J.: National Assessment of Educational Progress (NAEP), Educational Testing Service.

Linn, M. C. 1988, May. *Trends in the magnitude and nature of cognitive gender differences: Implications for the SAT.* Paper presented at ETS Seminar, Princeton, N.J.

Lupkowski, A. E. 1987, April. *Sex differences on the Differential Aptitude Test.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

McConeghy, J. I. 1987, April. *Mathematics attitudes and achievements: Gender differences in a multivariate context.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C. ERIC Document Reproduction Service No. ED 284 742.

Mullis, I. V. S. 1987, April. *Trends in performance for women taking the NAEP reading and writing assessment.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

National Assessment of Educational Progress. 1983. *The Third National Mathematics Assessments: Results, trends and issues.* Denver, CO: Education Commission of the States.

National Assessment of Educational Progress. 1985. *The Reading Report Card. Progress toward excellence in our schools: Trends in reading over four national assessments, 1971-1984.* ETS Report No. 15-R-01. Princeton, N.J.: Educational Testing Service.

National Assessment of Educational Progress. 1986. NAEP 1986 Mathematics Assessment, Weighted W.A.R.M. Background Factor Percentages and Mean Math Proficiency Composites. Unpublished raw data.

Ramist, L., and Arbeiter, S. 1986. *Profiles, college-bound seniors, 1985.* New York: College Entrance Examination Board.

Rock, D. A., Goertz, M. E., Ekstrom, R. B., Hilton, T. L., and Pollack, J. 1984, December. *Factors associated with test score decline.* Briefing paper. Princeton, N.J.: Educational Testing Service.

Rock, D. A., Hilton, T. L., Pollack, J., Ekstrom, R. B., and Goertz, M. E. 1985. *Psychometric analysis of the NLS and the High School and Beyond Test Batteries.* NCES Report No. 85-218. Washington, D.C.: National Center for Education Statistics.

Senk, S., and Usiskin, Z. 1983. Geometry proof writing: A new view of sex differences in mathematics ability. *American Journal of Education* 91:187-201.

Stanley, J. C. 1982, March. Identification of intellectual talent. In W. B. Schrader, ed., *New directions for testing and measurement: Measurement, guidance, and program improvement,* No. 13. San Francisco: Jossey-Bass.

Stanley, J. 1987, April. *Sex differences on the College Board Achievement Tests and the Advanced Placement Examinations.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Wheeler, P., and Harris, A. 1981. *Comparison of male and female performance on the ATP Physics Test.* CB Report No. 81-4. Princeton, N.J.: Educational Testing Service.

Wild, C. L. 1981. *A summary of data collected from Graduate Record Examinations test-takers during 1979-80.* ETS Data Summary Report No. 5. Princeton, N.J.: Educational Testing Service.

Wilder, G., Casserly, P., and Burton, N. 1988. *Young SAT-takers: Two surveys.* College Board Report No. 88-1. New York: College Entrance Examination Board.

Zerega, M. E., Haertel, G. D., Tsai, S.-L., and Waldberg, H. J. 1986. Late adolescent sex differences in science learning. *Science Education* 70(4):447-60.

Journal Articles and Research Reports

Literature and Research Reviews

Benbow, C. P. 1988. Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences,* in press.

Block, J. H. 1976. Issues, problems, and pitfalls in assessing sex differences: A critical review of *The Psychology of Sex Differences. Merrill-Palmer Quarterly* 22(4):283-308.

Chipman, S. F. 1988, March/April. Far too sexy a topic [Review of *The psychology of gender differences: Advances through meta-analysis*]. *Educational Researcher,* 46-49.

Deaux, K. 1985. Sex and gender. *Annual Review of Psychology* 36:49-81.

Farmer, H. S. 1987, March. A multivariate model for explaining gender differences in career and achievement motivation. *Educational Researcher,* 5-9.

Fennema, E. 1974. Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education* 5:126-29.

Levine, D. U., and Ornstein, A. C. 1983. Sex differences in ability and achievement. *Journal of Research and Development in Education* 16(2):66-72.

Lockheed, M. E., Thorpe, M., Brooks-Gunn, J., Casserly, P., and McAloon, A. 1985. *Sex and ethnic differences in middle school mathematics, science and computer science: What do we know?* A report submitted to The Ford Foundation. Princeton, N.J.: Educational Testing Service.

Tittle, C. K. 1986. Gender research and education. *American Psychologist* 41(10):1161-68.

Wharton, Y. L. 1977. *List of hypotheses advanced to explain the SAT decline.* New York: College Entrance Examination Board.

Psychology of Gender

Bleier, R. 1987. Science and belief: A polemic on sex differences research. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 111-30. Indianapolis: Indiana University Press.

Deaux, K., and Major, B. 1987. Putting gender into context: An interactive model of gender-related behavior. *Psychological Review* 94(3):369-89.

Gilligan, C. 1987. Remapping development: The power of divergent data. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 77-94. Indianapolis: Indiana University Press.

Jacklin, C. N. 1987. Feminist research and psychology. In C. Farnham, ed., *The impact of feminist research in the academy,* pp. 94-107. Indianapolis: Indiana University Press.

Wittig, M. A. 1985. Metatheoretical dilemmas in the psychology of gender. *American Psychologist* 40(7):800-11.

Meta-Analyses

Becker, B. J., and Hedges, L. V. 1984. Meta-analysis of cognitive gender differences: A comment on an analysis by Rosenthal and Rubin. *Journal of Educational Psychology* 76(4):583-87.

Hyde, J. S. 1981. How large are cognitive gender differences? A meta-analysis using w^2 and d. *American Psychologist* 36(8):892-901.

Hyde, J. S., and Linn, M. C. 1988, in press. A meta-analysis of gender differences in verbal abilities.

Meehan, A. M. 1984. A meta-analysis of sex differences in formal operational thought. *Child Development* 55:1110-24.

Rosenthal, R., and Rubin, D. B. 1982. Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology* 74(5):708-12.

Steinkamp, M. W., and Maehr, M. L. 1984. Gender differences in motivational orientations toward achievement in school science: A quantitative synthesis. *American Educational Research Journal* 21(1):39-59.

Sex Inequities in Education: Evidence, Causes and Potential Solutions

Dix, L. S. 1987. *Women: Their underrepresentation and career differentials in science and engineering.* Proceedings of a workshop. Washington, D.C.: National Academy Press, Office of Scientific and Engineering Personnel.

Kahle, J. B. 1984. *Girls in school/women in science: A Synopsis.* Paper presented at the annual Women's Studies Conference, Greeley, Colo. ERIC Document Reproduction Service No. ED 243 785.

Kahle, J. B., and Lakes, M. K. 1983. The myth of equality in science classrooms. *Journal of Research in Science Teaching* 20(2):131-40.

Leinhardt, G., Seewald, A. M., and Engel, M. 1979. Learning what's taught: Sex differences in instruction. *Journal of Educational Psychology* 71(4):432-39.

Lockheed, M. E. 1984. Sex segregation and male preeminence in elementary classrooms. In E. Fennema and M. J. Ayer, eds., *Women and education: Equity or equality?* pp. 117-35. Berkeley, Calif.: McCutchan.

Peterson, P. L., and Fennema, E. 1985. Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal* 22(3):309-35.

Sadker, M., and Sadker, D. 1985a, January. Is the O.K.

classroom O.K.? *Phi Delta Kappan.*
Sadker, M., and Sadker, D. 1985b, March. Sexism in the schoolroom of the 80's. *Psychology Today* 54–57.
Sadker, M., and Sadker, D. 1986, March. Sexism in the classroom: From grade school to graduate school. *Phi Beta Kappan.*

Characteristics of Tests

American College Testing Program. 1988. *ACT Assessment Program Technical Manual.* Iowa City, Iowa: ACT.
Breland, H. M., and Griswold, P. A. 1982. Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology* 74(5):713–21.
Bridgeman, B. 1988. *Comparative validity of multiple-choice and free-response advanced placement biology items.* Research report draft, submitted for review. Princeton, N.J.: Educational Testing Service.
Burton, N. 1988, April. *Modeling women's performance on the SAT.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Carlton, S. T. 1987, July. *Differences in male and female performance on standardized verbal tests.* Paper presented at the Third International Interdisciplinary Congress on Women, Dublin, Ireland.
Chipman, S. F. 1988, April. *Word problems: Where test bias creeps in.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Cohen, J. 1977. *Statistical power analysis for the behavioral sciences.* Revised ed. New York: Academic Press.
Donlon, T. F., ed. 1984. *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests.* New York: College Entrance Examination Board.
Donlon, T. F., Hicks, M. M., and Wallmark, M. M. 1980. Sex differences in item responses on the Graduate Record Examination. *Applied Psychological Measurement* 4(1):9–20.
Doolittle, A. E. 1985, April. *Understanding differential item performance as a consequence of gender differences in academic background.* Paper presented at the annual meeting of the American Education Research Association, Chicago, Ill.
Doolittle, A. E. 1987, August. *Gender differences in performance on mathematics achievement items.* Paper presented at the annual meeting of the American Psychological Association, New York.
Doolittle, A. E., and Cleary, T. A. 1987. Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement* 24(2)157–66.
Dorans, N. J. 1982. *Technical review of SAT item fairness studies: 1975–1979.* ETS Statistical Report No. SR-82-90. Princeton, N.J.: Educational Testing Service.
Grandy, J. 1987a, October. *Ten-year trends in SAT scores and other characteristics of high school seniors taking the SAT and planning to study mathematics, science, or engineering.* ETS Research Report No. 87-49. Princeton, N.J.: Educational Testing Service.
Grandy, J. 1987b, October. *Trends in the selection of science, mathematics, or engineering as major fields of study among top-scoring SAT takers.* ETS Research Report No. 87-39. Princeton, N.J.: Educational Testing Service.
Grandy, J., and Courtney, R. 1985. *Factors contributing to the changing characteristics of prospective humanities majors: 1975–1984.* Grant No. OP-20193-84. Princeton, N.J.: Educational Testing Service.
Hudson, L. 1986. Item-level analysis of sex differences in mathematics achievement test performance. *Dissertation Abstracts International* 47(2): order no. DA8607283.
Jones, R. F., and Vanyur S. 1985, April. *An investigation of gender-related test bias for the Medical College Admission Test.* Paper presented at the meeting of the National Council on Measurement in Education, Chicago, Ill.
Lawrence, I. M., Curley, W. E., and McHale, F. J. 1987. *Differential item functioning of SAT-Verbal reading subscore items for male and female examinees.* ETS Research Report, in press. Princeton, N.J.: Educational Testing Service.
Linn, M. C., De Benedictis, T., Delucchi, K., Harris, A., and Stage, E. 1987. Gender differences in National Assessment of Educational Progress science items: What does "I don't know" really mean? *Journal of Research in Science Teaching* 24(3):267–78.
Linn, R. L. 1982. Ability testing: Individual differences, predictions and differential prediction. In A. Wigdor and W. Garner, eds., *Ability testing: Uses, consequences and controversies,* pp. 335–38. Washington, D.C.: National Academy Press.
Linn, M. C., and Hyde, J. S. 1988, April. *Gender differences in verbal ability: A meta-analysis.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Loewen, J. W., Rosser, P., and Katzman, J. 1988, April. *Gender bias in SAT items.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Lubetkin, J. 1988, April. *The Scholastic Aptitude Test: A valid and unbiased predictor of college performance?* Unpublished B.A. Thesis, Princeton University, Princeton, N.J.
McPeek, W. M., and Wild, C. L. 1987, April. *Identifying differentially functioning items in the NTE core battery.* Unpublished research report. Princeton, N.J.: Educational Testing Service.
McPeek, W. M., and Wild, C. L. 1987, August. *Characteristics of quantitative items that function differently for men and women.* Paper presented at the annual meeting of the American Psychological Association, New York.
Murphy, R. J. L. 1982. Sex differences in objective test performance. *British Journal of Educational Psychology* 52:213–19.
Pearlman, M. A. 1987, April. *Trends in women's total score and item performance on verbal measures: Five forms of the GRE: Verbal items that display large Mantel-Haenszel DIF values.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.
Pennock-Roman, M., Rock, D. A., and Enright, M. K. 1988, January. *Language background and test validity for Hispanic-American students. Part I: Comparisons between Hispanic and non-Hispanic-White groups.* Unpublished

manuscript. Princeton, N.J.: Educational Testing Service.

Ramist, L. 1984. Predictive validity of the ATP tests. In T. Donlon, ed., *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests.* New York: The College Board.

Swinton, S. S. 1987. *The predictive validity of the restructured GRE with particular attention to older students.* GRE Board Professional Report No. 83-25P, ETS RR No. 87-22. Princeton, N.J.: Educational Testing Service.

Welch, C. J., and Doolittle, A. E. 1988, April. *Gender-based Differential Item Performance in English usage items.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Wendler, C. L. W., and Carlton, S. T. 1987, April. *An examination of SAT verbal items for differential performance by women and men: An exploratory study.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Wild, C. L., and Dwyer, C. A. 1980. Sex bias in selection. In L. J. Th. van der Kamp, W. F. Langerak, and D. N. M. de Gruijter, eds., *Psychometrics for educational debates,* pp. 153-68. New York: Wiley.

Wild, C. L., and McPeek, W. M. 1986, August. *Performance of the Mantel-Haenszel Statistic in identifying differentially functioning items.* Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Wittig, M. A., Sasse, S. H., and Giacomi, J. 1984. Predictive validity of five cognitive skills tests among women receiving engineering training. *Journal of Research in Science Teaching* 21(5):537-46.

Population and Demographic Trends

Burton, N. W. 1988, April. *Modeling women's performance on the SAT.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Burton, N. W., Lewis, C. and Robertson, N. 1988, April. Draft. *SAT gender differences controlled for population trends.* Princeton, N.J.: Educational Testing Service.

Grant, C. A., and Sleeter, C. E. 1986. Race, class, and gender in education research: An argument for integrative analysis. *Review of Educational Research* 56(2):195-211.

Paulhus, D., and Shaffer, D. R. 1981. Sex differences in the impact of older and number of younger siblings on scholastic aptitude. *Social Psychology Quarterly* 44:363-68.

Steelman, L. C., and Mercy, J. A. 1983. Sex differences in the impact of the number of older and younger siblings on IQ performance. *Social Psychology Quarterly* 46(2):157-62.

Zajonc, R. B. 1986. The decline and rise of scholastic aptitude scores: A prediction derived from the confluence model. *American Psychologist* 41(8):862-67.

Zajonc, R. B., and Bargh, J. 1980a. Birth order, family size and decline of SAT scores. *American Psychologist* 35:662-68.

Zajonc, R. B., and Bargh, J. 1980b. The confluence model: Parameter estimation for six divergent data sets on family factors and intelligence. *Intelligence* 4:349-61.

Differential Coursework, Participation, and Enrollment

Armstrong, J. M. 1985. A national assessment of participation and achievement of women in mathematics. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds. *Women and mathematics: Balancing the equation.* Hillsdale, N.J.: Lawrence Erlbaum.

Jones, L. V., Davenport, E. C., Bryson, A., Bekhuis, T., and Zwick, R. 1986. Mathematics and science test scores as related to courses taken in high school and other factors. *Journal of Educational Measurement* 23(3):197-208.

Laing, J., Engen, H., and Maxey, J. 1987. *Relationships between ACT test scores and high school courses.* Research report. Iowa City, Iowa: American College Testing Program.

Noble, J., and McNabb, T. 1988, April. *Differential coursework in high school: Implications for performance on the ACT assessment.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Pallas, A. M., and Alexander, K. L. 1983. Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. *American Educational Research Journal* 20(2):165-82.

Stallings, S. J. 1985. School, classroom and home influences on women's decisions to enroll in advanced mathematics courses. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds. *Women and mathematics: Balancing the equation.* Hillsdale, N.J.: Lawrence Erlbaum.

Wise, L. L. 1985. Project TALENT: Mathematics course participation in the 1960s and its career consequences. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds. *Women and mathematics: Balancing the equation,* pp. 25-58. Hillsdale, N.J.: Lawrence Erlbaum.

Cognitive and Learning Styles

Ash, B. F. 1986. *Identifying learning styles and matching strategies for teaching and learning.* ERIC Document Reproduction Service No. ED 270 142.

Cox, P. W., and Witkin, H. A. 1978. *Field dependence-independence and psychological differentiation: Bibliography with index, Supplement No. 3.* ETS Research Bulletin No. 78-8. Princeton, N.J.: Educational Testing Service.

Goodenough, D. R., and Witkin, H. A. 1977. *Origins of field-dependent and field-independent cognitive styles.* ETS Research Bulletin No. 77-9. ERIC Document Reproduction Service No. ED 150 155. Princeton, N.J.: Educational Testing Service.

Huber, G. L. 1988, April. *Preference for learning situations and uncertainty orientation: A cross-cultural comparison.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Messick, S. 1976. Personality consistencies in cognition and creativity. In S. Messick, ed., *Individuality in learning: Implications of cognitive style and creativity for human development,* pp. 4-22. San Francisco: Jossey-Bass.

Messick, S. 1984. The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist* 19(2):59-74.

Nisbett, R. E., and Wilson, T. D. 1977. Telling more than

we can know: Verbal reports on mental processes. *Psychological Review* 84:231–59.
Rutter, M. 1977. Individual differences. In M. Rutter and L. Hersov, eds, *Child psychiatry: Modern approaches.* Oxford, England: Blackwell Scientific.
Slavin, R. E. 1978. Effects of student teams and peer tutoring on academic achievement and time-on-task. *Journal of Experimental Education* 48:252–57.
Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. G., and Karp, S. A. 1962. *Psychological differentiation.* New York: Wiley.
Witkin, H. A., and Goodenough, D. R. 1981. Cognitive styles: Essence and origins. *Psychological Issues.* Monograph 51. New York: International Universities Press.
Witkin, H. A., Goodenough, D. R., and Oltman, P. K. 1979. Psychological differentiation: Current status. *Journal of Personality and Social Psychology* 37(7):1127–45.

Sex-Role Socialization

Block, J. H. 1983. Differential premises arising from differential socialization of the sexes: Some conjectures. *Child Development* 54:1335–54.
Boswell, S. L. 1985. The influence of sex-role stereotyping on women's attitudes and achievement in mathematics. In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds, *Women and mathematics: Balancing the equation,* pp. 175–197. Hillsdale, N.J.: Lawrence Erlbaum.
Cherry, L., and Lewis, M. 1975. Mothers and two-year-olds: A study of sex-differentiated aspects of verbal interaction. *Developmental Psychology* 12(4):278–82.
Fagot, B. I. 1978. The influence of sex of child on parental reactions to toddler children. *Child Development* 49:459–65.
Grieb, A., and Easley, J. 1984. A primary school impediment to mathematics equity: Case studies in rule-dependent socialization. In M. W. Steinkamp and M. L. Maehr, eds, *Advances in motivation and achievement. Vol. 2: Women in science.* Greenwich, Conn.: Jai Press.
Horner, M. S. 1970. Femininity and successful achievement: A basic inconsistency. In J. M. Bardwick, et al., eds., *Feminine personality and conflict.* Monterey, Calif.: Brooks/Cole.
Houston, A. C. 1983. Sex-typing. In P. H. Mussen, ed., *Handbook of child psychology,* vol. 4. New York: Wiley.
Jacobs, J. E., and Eccles, J. S. 1985, March. Gender differences in math ability: The impact of media reports on parents. *Educational Researcher,* 20–25.
Kagan, J. 1964. The acquisition and significance of sex-typing and sex-role identity. In M. Hoffman and L. Hoffman, eds., *Review of child development research,* vol. 1. New York: Russell Sage.
Kavrell, S. M., and Peterson, A. C. 1984. Patterns of achievement in early adolescence. In M. W. Steinkamp and M. L. Maehr, eds., *Advances in motivation and achievement.* Vol. 2: Women in science. Greenwich, Conn.: Jai Press.
Lewis, M., and Freedle, R. 1973. The mother-infant dyad. In P. Pliner, L. Kranes, and T. Alloway, eds., *Communication and affect: Language and thought.* New York: Academic Press.
Milton, G. A. 1957. The effects of sex-role identification upon problem-solving skill. *Journal of Abnormal and Social Psychology* 55:208–12.
Raymond, C. L., and Benbow, C. P. 1986. Gender differences in mathematics: A function of parental support and student sex typing? *Developmental Psychology* 22(6):808–19.
Rheingold, H. L., and Cook, K. V. 1975. The content of boys' and girls' rooms as an index of parent behavior. *Child Development* 46:459–63.
Schickedanz, J. A. 1973. The relationship of sex-typing of reading to reading achievement and reading choice behavior in elementary school boys. *Dissertation Abstracts* 34(12A Pt. 1):7645.
Women on Words and Images. 1972. *Dick and Jane as Victims: Sex stereotyping in children's readers.* Princeton, N.J.: Women on Words and Images.

Attitudes, Expectations, and Motivation

Condry, J., and Dyer, S. 1976. Fear of success: Attributions of cause to the victim. *Journal of Social Issues* 32:63–83.
Diener and Dweck, C. S. 1980. An analysis of learned helplessness: The processing of success. *Journal of Personality and Social Psychology,* 39:940–50.
Dweck, C. S., Davidson, W., Nelson, S., and Enna, B. 1978. Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental Psychology* 14(3):268–76.
Eccles (Parsons), J. 1983. Expectancies, values, and academic behaviors. In J. T. Spence, ed., *Achievement and achievement motives: Psychological and sociological approaches.* San Francisco: Freeman.
Eccles, J. S. 1985. Sex differences in achievement patterns. In T. Sonderegger, ed., *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.
Eccles, J. S. 1986. Gender-roles and women's achievement. *Educational Researcher* 15:15–19.
Eccles, J. S. 1987. Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly* 11:135–72.
Eccles (Parsons), J., Adler, T., and Meece, J. L. 1984. Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology* 46(1):26–43.
Ethington, C. A., and Wolfle, L. M. 1986. A structural model of mathematics achievement for men and women. *American Educational Research Journal,* 23(1):65–75.
Frieze, I. H. 1980. Beliefs about success and failure in the classroom. In J. McMillan, ed., *The social psychology of school learning.* New York: Academic Press.
Frieze, I. H., Whitley, B. E., Hanusa, B. H., and McHugh, M. 1982. Assessing the theoretical models for sex differences in causal attributions for success and failure. *Sex Roles* 8:333–45.
Haertel, G. D., Walberg, H. J., Junker, L., and Pascarella, E. T. 1981. Early adolescent sex differences in science learning: Evidence from the National Assessment of Educational Progress. *American Educational Research Journal* 18(3):329–41.
Karmos, A. H., and Karmos, J. S. 1984, July. Attitudes toward standardized achievement tests and their rela-

tion to achievement test performance. *Measurement and Evaluation in Counseling and Development,* 56–66.
Lenny, E. 1977. Women's self-confidence in achievement settings. *Psychological Bulletin* 84:1–13.
Lepper, M., 1985. Microcomputers in education: Motivational and social issues. *American Psychologist* 40:1–18.
Licht, B. G., and Dweck, C. S. 1983. Sex differences in achievement orientations: Consequences for academic choices and attainments. In M. Marland, ed., *Sex differentiation and schooling.* London: Heinemann.
Weiner, B. 1979. A theory of motivation for some classroom experiences. *Journal of Educational Psychology* 71:3–25.
Weiner, B., Frieze, I. H., Kukla, A., Reed, L., Rest, S., and Rosenbaum, R. M. 1971. *Perceiving the causes of success and failure.* Morristown, N.J. General Learning Press.
Whitley, B. E., Jr., and Frieze, I. H. 1985. *The effect of question wording style and research context on attributions for success and failure: A meta-analysis.* Paper presented at the annual meeting of the Eastern Psychological Association, Boston.
Wolleat, P. L., Pedro, J. D., Becker, A. D., and Fennema, E. 1980. Sex differences in high school students' causal attributions of performance in mathematics. *Journal for Research in Mathematics Education* 11:356–66.

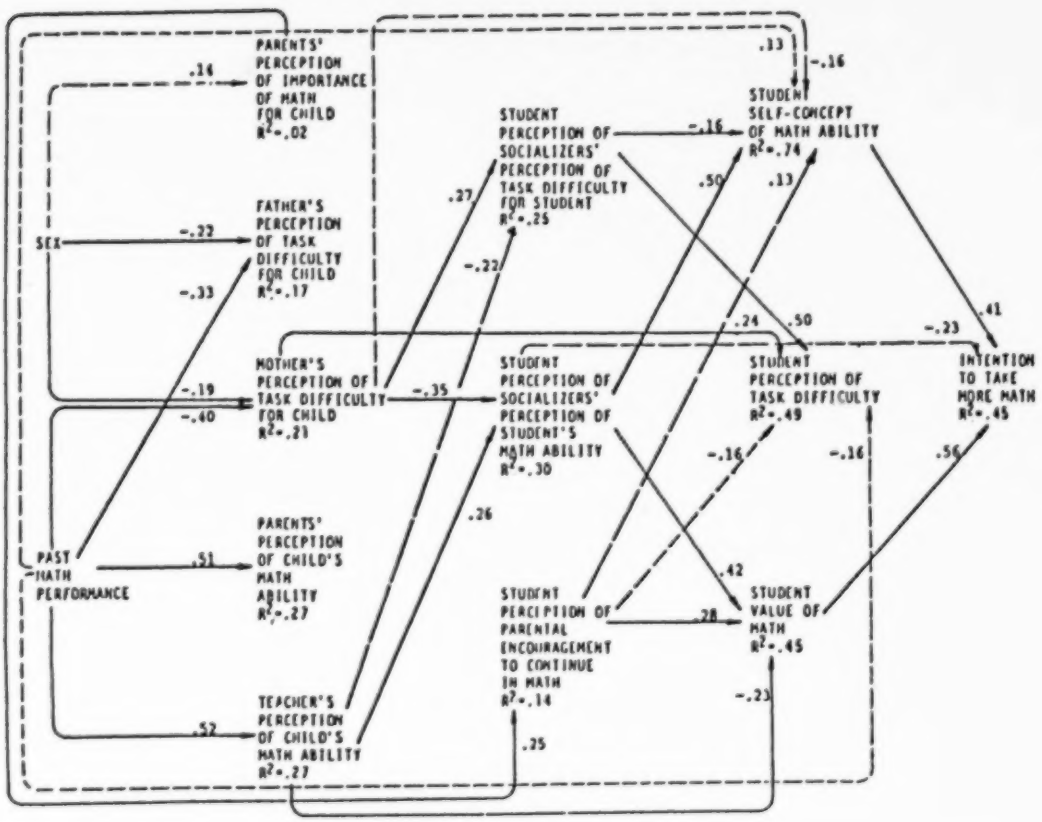
Biological Sex Differences

Annett, M. 1980. Sex differences in laterality—meaningfulness vs. reliability. *The Behavioral and Brain Sciences* 3:227–63.
Benbow, C. P. 1986. Physiological correlates of extreme intellectual precocity. *Neuropsychologia* 24:719–25.
Benbow, C. P., Benbow, R. M. 1984. Biological correlates of high mathematical reasoning ability. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 469–90. New York: Elsevier.
Butler, S. 1984. Sex differences in human cerebral function. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 443–55. New York: Elsevier.
Dunn, B. R. 1988, April. *Gender differences in EEG patterns: Are they indexes of different cognitive strategies.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.
Heister, G. 1984. Sex differences in visual half-field superiority as a function of responding hand and motor demands. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 457–68. New York: Elsevier.
Kimura, D., and Harshman, R. A. 1984. Sex differences in brain organization for verbal and non-verbal functions. In G. J. De Vries, J. P. C. De Bruin, H. B. M. Uylings, and M. A. Corner, eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 423–41. New York: Elsevier.
Lehrke, R. G. 1974. *X-linked mental retardation and verbal disability.* New York: Intercontinental Medical Book.
McGlone, J. 1980. Sex differences in human brain asymmetry: A critical survey. *The Behavioral and Brain Sciences* 3:215–63.
Newcombe, N., and Dubas, J. S. 1987. Individual differences in cognitive ability: Are they related to timing of puberty? In R. M. Lerner and T. L. Foche, ed., *Biological-psychosocial interactions in early adolescence.* pp. 249–302. Hillsdale, N.J.: Lawrence Erlbaum Associates.
Nyborg, H. 1984. Performance and intelligence in hormonally different groups. In G. J. De Vries, et al., eds., *Progress in brain research. Vol. 61: Sex differences in the brain,* pp. 491–508. New York: Elsevier.
Peterson, A. C. 1983. Pubertal change and cognition. In J. Brooks-Gunn and A. C. Peterson, eds., *Girls at puberty: Biological and psychosocial perspectives,* pp. 179–98. New York: Plenum Press.
Stafford, R. E. 1972. Hereditary and environmental components of quantitative reasoning. *Review of Educational Research* 42:183–201.
Vandenberg, S. G. 1968. Primary mental abilities or general intelligence? Evidence from twin studies. In J. M. Thoday and A. S. Parkes, eds., *Genetic and environmental influences on behavior,* pp. 146–60. New York: Plenum.
Waber, D. P., Mann, M. B., Merola, J., and Moylan, P. M. 1985. Physical maturation rate and cognitive performance in early adolescence: A longitudinal examination. *Developmental Psychology* 21(4):666–81.

Spatial-Abilities Research

Ben-Chaim, D., Lappan, G., and Houang, R. T. 1988. The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal* 25(1):51–71.
Caplan, P. J., MacPherson, G. M., and Tobin, P. 1985. Do sex-related differences in spatial abilities exist? *American Psychologist* 40(7): 786–99.
Connor, J. M., and Serbin, L. A. 1985. Visual-spatial skill: Is it important for mathematics? Can it be taught? In S. F. Chipman, L. R. Brush, and D. M. Wilson, eds., *Women and mathematics: Balancing the equation,* pp. 151–74. Hillsdale, N.J.: Lawrence Erlbaum.
Crosson, C. W. 1984. Age and field independence among women. *Experimental Aging Research* 10:165–70.
Fennema, E., and Tartre, L. A. 1985. The use of spatial visualization in mathematics by girls and boys. *Journal for Research in Mathematics Education* 16:(3)184–206.
Hyde, J. S., Geiringer, E. R., and Yen, W. M. 1975. On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. *Multivariate Behavioral Research* 10:289–309.
Linn, M. C., and Peterson, A. C. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development* 56:1479–98.
Linn, M. C., and Peterson, A. C. 1986. A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde and M. C. Linn, eds. *The Psychology of gender: Advances through meta-analysis,* pp. 67–101. Baltimore, Md.: Johns Hopkins University Press.
Sherman, J. A. 1967. Problems of sex differences in space perception and aspects of intellectual functioning. *Psychological Review* 74:290–99.

APPENDIX B. SELECTED MODELS OF INFLUENCES ON GENDER-BASED DIFFERENCES



Note: Dashed lines are significant at $p \leq .05$, solid lines at $p \leq .01$; N = 1.56.
Standardized beta weights are shown on path.
R^2 = percent of variance accounted for on each criterion measure by all preceding predictor variables; each R^2 is listed under its criterion measure.

Figure 1. Reduced path-analytic diagram for test of socialization model.

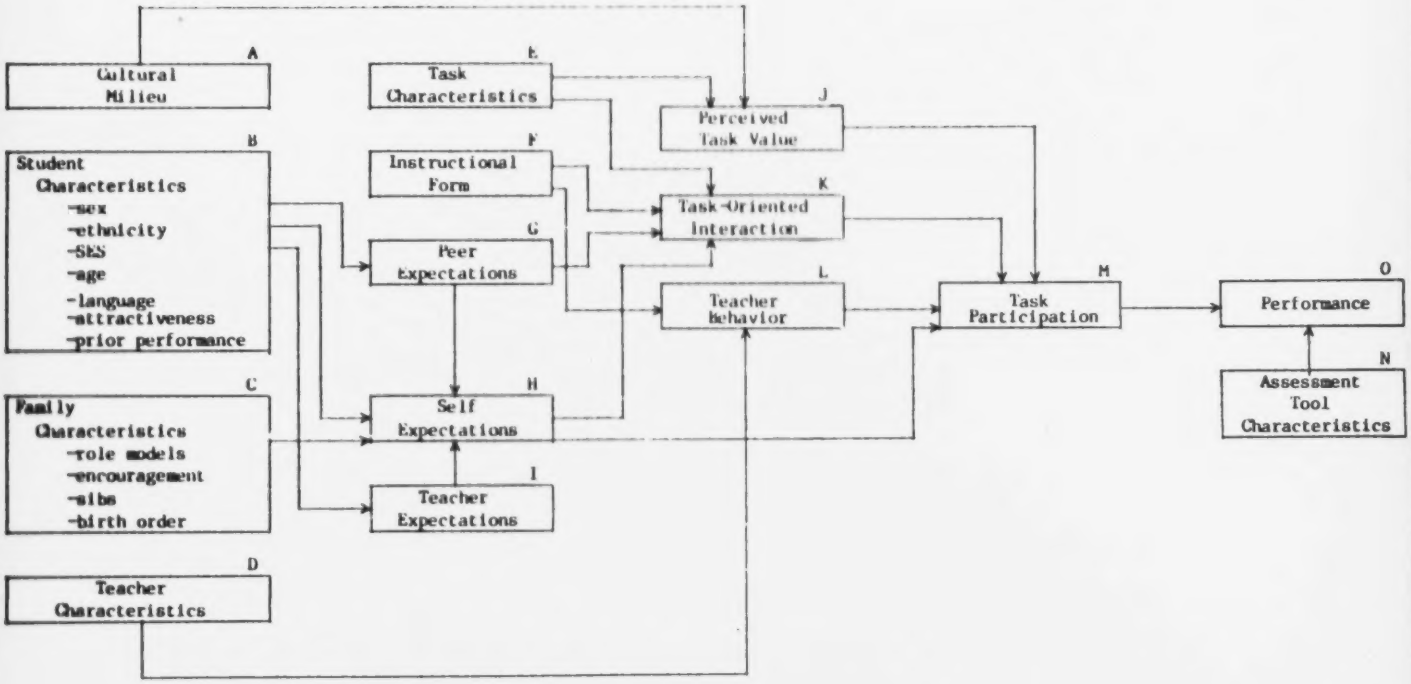


Figure 2. Task-performance model of mathematics, science, or computer performance.

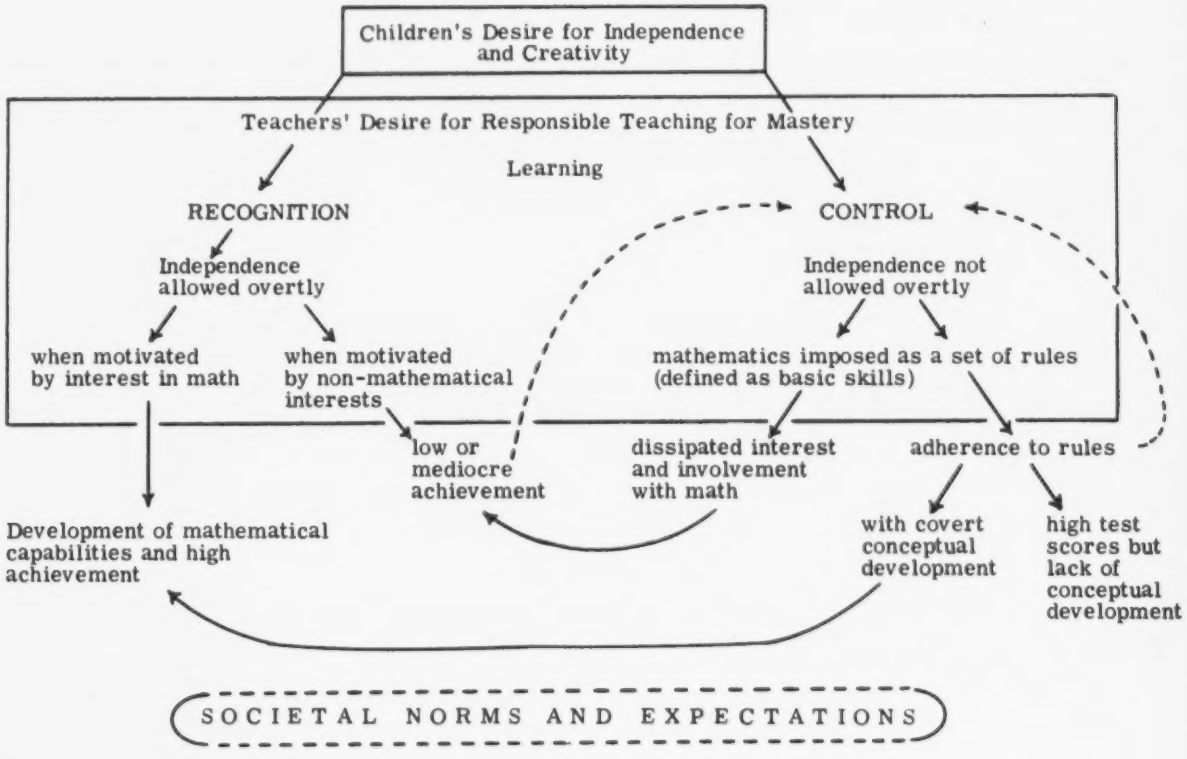
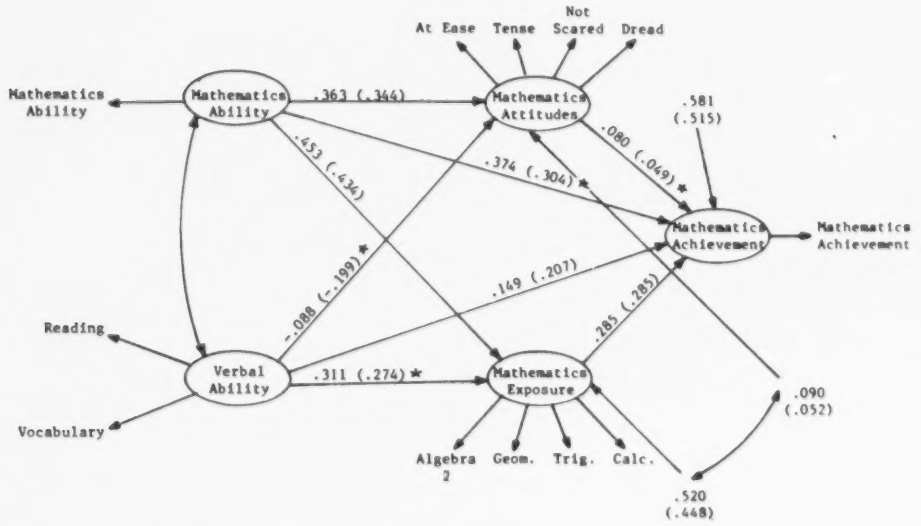

Children's Desire for Independence and Creativity
Teachers' Desire for Responsible Teaching for Mastery
Learning
RECOGNITION
CONTROL
Independence allowed overtly
Independence not allowed overtly
when motivated by interest in math
when motivated by non-mathematical interests
mathematics imposed as a set of rules (defined as basic skills)
low or mediocre achievement
dissipated interest and involvement with math
adherence to rules
with covert conceptual development
high test scores but lack of conceptual development
Development of mathematical capabilities and high achievement
SOCIETAL NORMS AND EXPECTATIONS

Figure 3. Alternative pathways of mathematical development.



Note: In each pair of coefficients, values for men are given first and values for women are given second (in parentheses). Pairs of coefficients found to be significantly different are marked with an asterisk. All coefficients are at least twice their standard errors. The numbers shown by residual error terms are coefficients of determination.

Figure 4. Structural equation and measurement models of mathematics achievement.

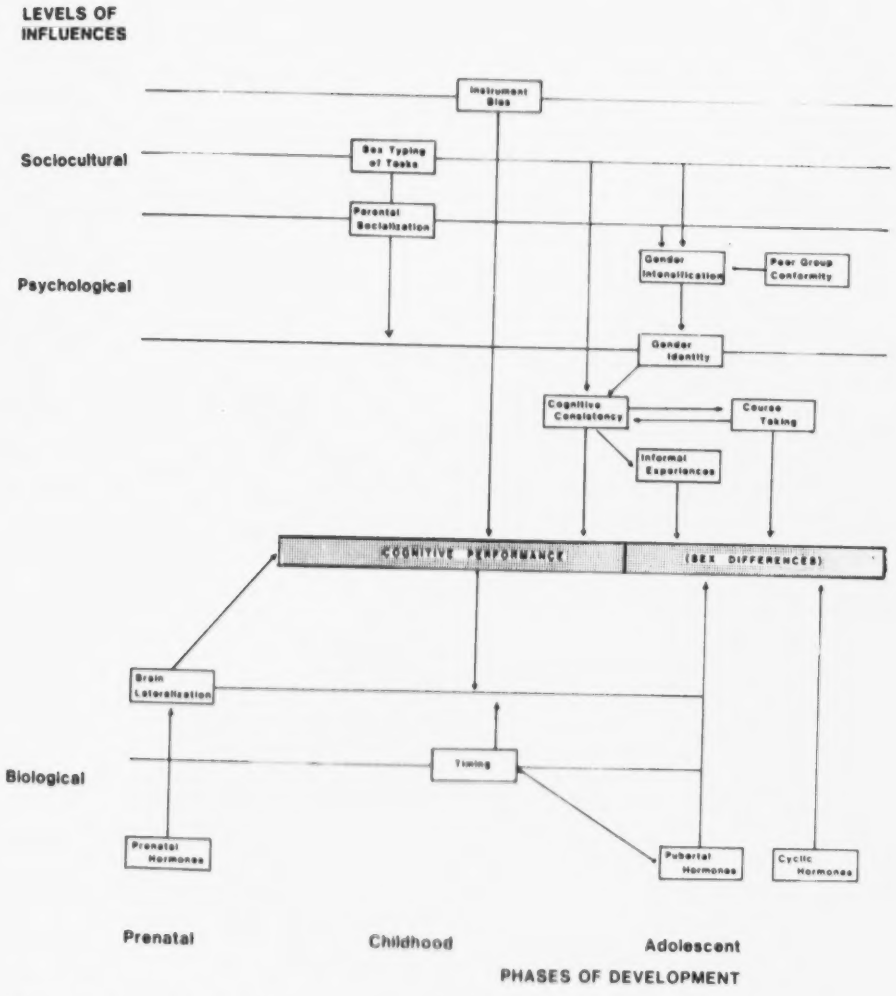
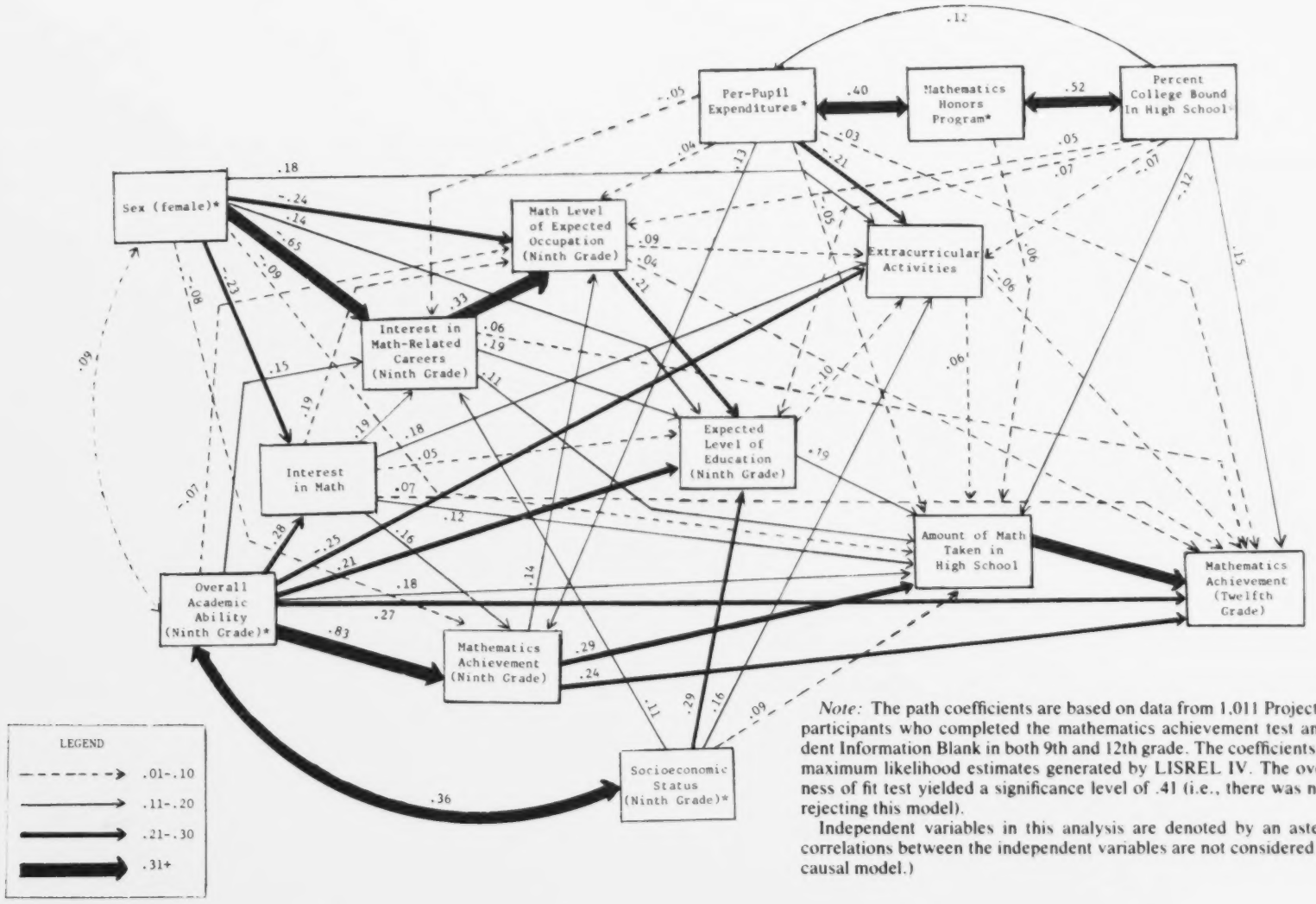


Figure 5. A model of biopsychosocial influences on cognitive performance.



Note: The path coefficients are based on data from 1,011 Project TALENT participants who completed the mathematics achievement test and the Student Information Blank in both 9th and 12th grade. The coefficients shown are maximum likelihood estimates generated by LISREL IV. The overall goodness of fit test yielded a significance level of .41 (i.e., there was no basis for rejecting this model).

Independent variables in this analysis are denoted by an asterisk. (The correlations between the independent variables are not considered part of the causal model.)

Figure 6. Summary path model of the relationship of sex to high school mathematics achievement.